

iEval as a Tool to Collect Student Evaluations of Teaching: Current Issues and Possible Improvements

Summary

This memo summarizes analyses of the survey used to collect student evaluations of teaching through iEval. The underlying assumption is that this process could be improved. Major findings include:

- iEval currently uses a five-point Likert scale but a seven-point scale would likely generate a wider distribution of responses.
- There are several items whose use or purpose are not clear, specifically those asking about student behaviors and the “blank” items in Section D. It may be advantageous to consider rewording items addressing information technology.
- A factor analysis of the current iEval questionnaire indicates that items group together into three scales: instructor behaviors, course materials, and student behaviors. The group of items related to instructor behaviors explains most of the overall variance. This suggests there are items that could be removed with little impact on the overall instrument.

Introduction

This memo reports on a set of analyses undertaken to examine the structure, content, and distribution of responses on the iEval survey currently used to collect student evaluation of teaching. These analyses are generally motivated by the sense that student evaluation of teaching as currently practiced at UCR could provide more useful information to improve teaching and learning on the campus. Specifically, there is relatively little variation in responses to the 19 Likert scale items with most instructors receiving a four or five on the five-point scale. It is possible that UCR instructors are, across the board, excellent. An alternative interpretation might be that variation in instructional quality is not accurately captured by the current instrument. Anecdotal evidence from faculty, departmental chairs, and administrators suggests the compression of answers near one end of the scale limits the quality and usefulness of the data provided. (A sample of the iEval questionnaire is provided in Appendix One.)

Another general observation is that shorter surveys are preferred to longer ones. Shorter surveys are likely to garner better response rates and may also encourage respondents to give more thought to each answer.

Number of Choices on Likert Scale

The relatively low level of variability across items may be driven by the use of a Likert scale with just five choices. With five choices students may not be presented with options that give them sufficiently detailed alternatives. Expanding the number of choices offered may increase clarity in the feedback students can provide.

iEval has used a seven point Likert scale in the past, and it is possible to compare the distribution of responses in historical data. iEval used a seven-point scale from the Spring 2005 to Spring 2006, and a five-point scale was used from Spring 2009 to Spring 2010. These time points offer four items that are similar (although not identical), and it is also possible to identify a subset of instructors who taught the same course at each time point. This produced a sample of 4,200-4,600 student responses on each item at each time point.

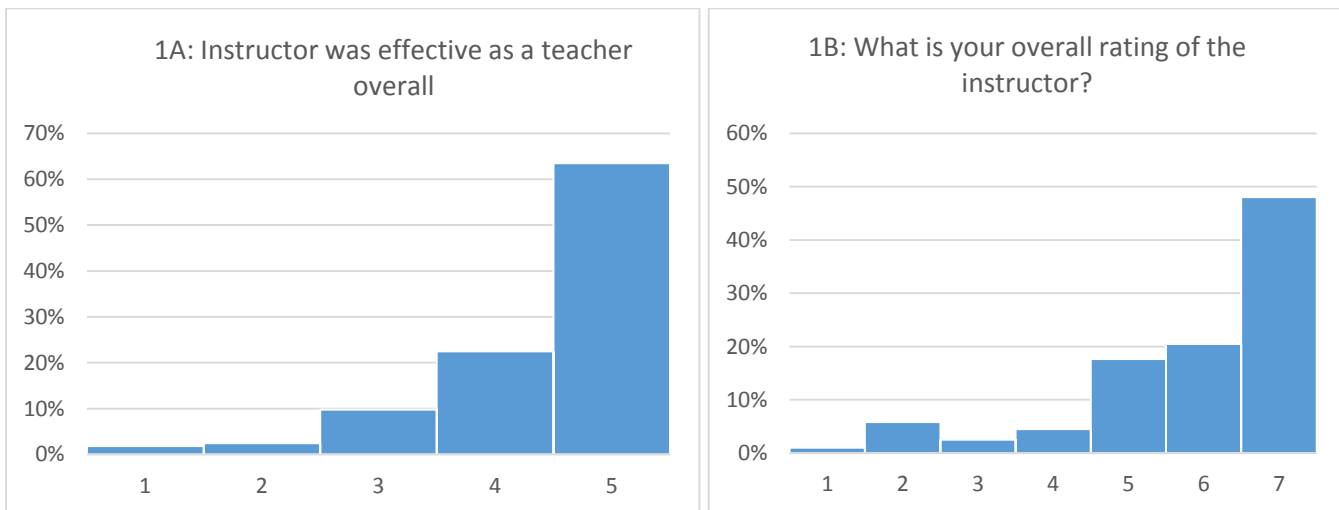
Table 1 shows a number of comparisons between this set of items on the five- and seven-point Likert scales. For each of the seven-point items the 25th, 50th and 75th percentiles have different values; for the five-point scale the observations at the 50th and 75th percentile have the same value. The standard deviations observed

for each of the seven-point items are significantly larger than for the five-point counterpart. The lower portion of Table 1 presents results from an F-test assessing the differences in variance in the group of responses collected with a five-point scale versus a seven-point scale. The variation is higher with the group of items on the seven-point scale; the F-test indicates the difference is significantly different from zero.

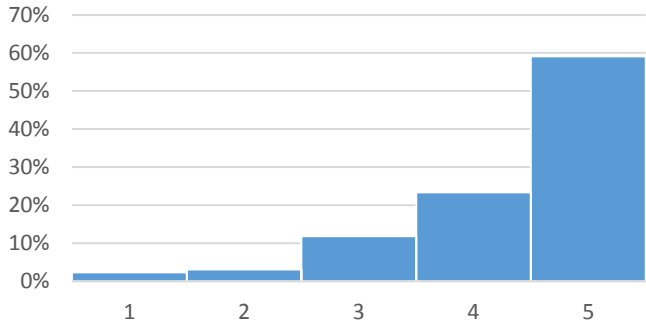
Table 1: Descriptive Statistics and F-test for Differences in Variance between Seven- and Five-Point Scales

Results	Seven	Five	Seven	Five	Seven	Five	Seven	Five
Mean	5.85	4.43	5.62	4.34	5.95	4.45	5.87	4.47
25 th Percentile	5	4	5	4	5	4	5	4
50 th Percentile (Median)	6	5	6	5	6	5	6	5
75 th Percentile	7	5	7	5	7	5	7	5
Variance	2.23	0.81	2.60	0.94	2.07	0.84	2.12	0.77
Standard Deviation	1.49	0.90	1.61	0.97	1.44	0.92	1.45	0.88
Observations	4399	4584	4392	4551	4384	4590	4179	4592
Degrees of Freedom	4398	4583	4391	4550	4383	4589	4178	4591
F statistics	2.76		2.76		2.46		2.73	
P(F<=f) one-tail	0.00		0.00		0.00		0.00	
F Critical one-tail	1.05		1.05		1.05		1.05	

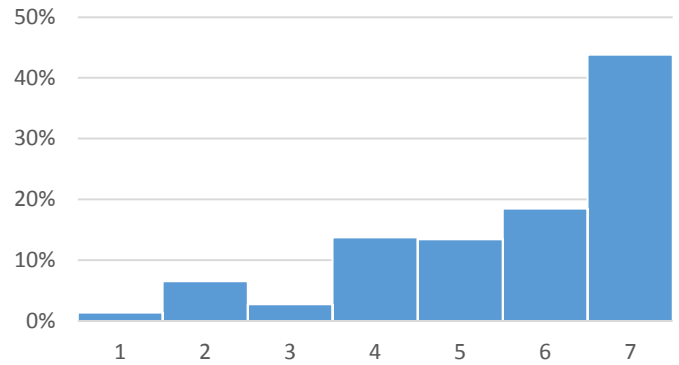
Figures 1A through 4B show histograms for responses to each item. One can see that most students select answers above the midpoint with both five- and seven-point scales. However, when a seven-point scale is used, this large group of responses is spread across three categories instead of just two. Also, on each of the four items examined here, there is a modest increase in the number of respondents who select an item below the midpoint. These analyses suggest that offering students a greater number of Likert choices may produce a more diverse distribution of responses.



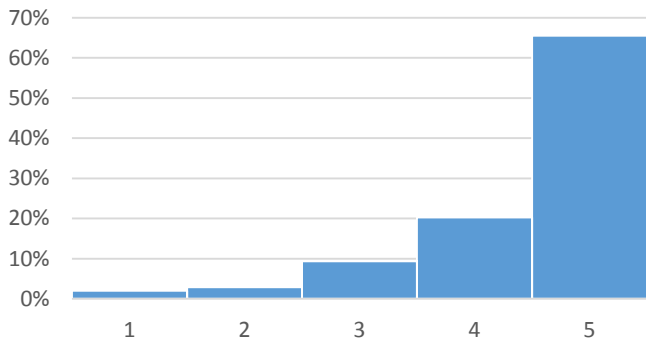
2A: The course overall as a learning experience was excellent



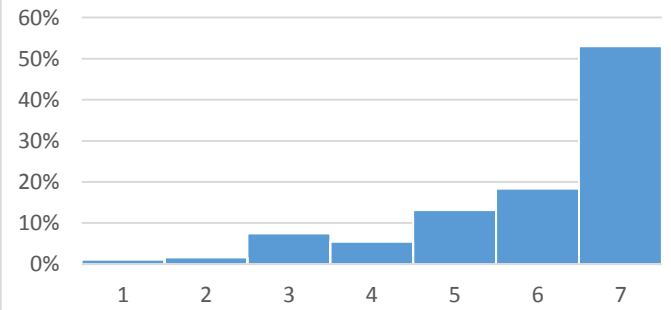
2B: What is your overall rating of the course?



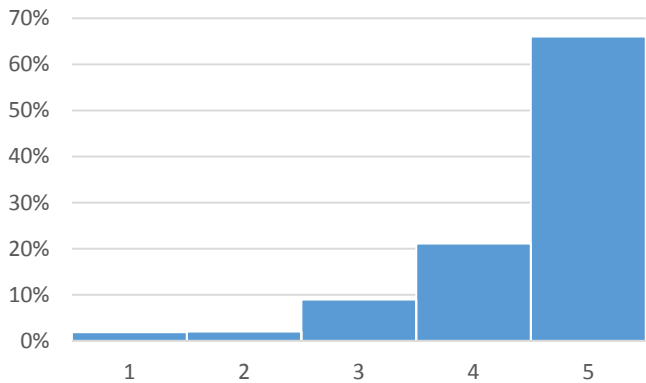
3A: Instructor was clear and understandable



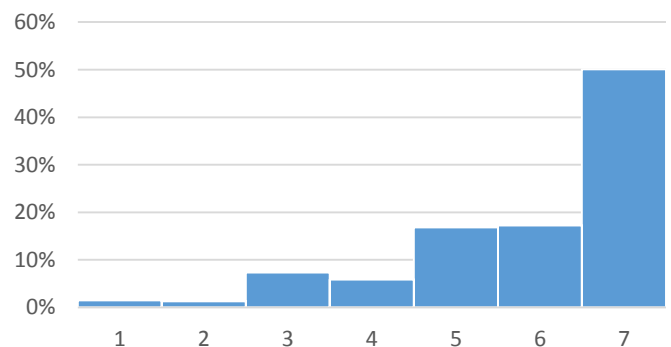
3B: Did the instructor present the material in an organized, understandable manner?



4A: Instructor respected students; sensitive to and concerned with their progress



4B: Was the instructor concerned about students learning and understanding the course material?



Content of Current Survey Items

Some of the items on the iEval questionnaire might be profitably reexamined. First, several items asking about instructor's behaviors link two ways to evaluate instructors. For example one item asks if the instructor was "available and helpful" and another if the instructor was "organized and prepared." While these concepts are related, best practice in survey construction suggest eliminating one of the attributes or splitting these into two items without the use of "and" would produce items that are conceptually clearer.

Second, the group of items on student behavior and attitudes would seem to give the ability to tease out patterns in responses based on student attitudes. (For example, are evaluations higher from students who were more motivated to take the course?) However, only class averages are made available to instructors which precludes this kind of analysis. If the data collected from these items is not being used it may be wise to consider removing them.

Third, there is only one item that asks about supplementary materials. Given rapid changes in instructional technology over the last few years this item now appears dated. It may be wise to consider revamping this one item to use wording more in-line with today's instructional technology.

Fourth, there is a group of "blank" items (in Section D) without text for questions but with a set of response categories. It is not clear how often these are used or for what purposes, although it would seem this format worked better when instructors could provide text for these items in the classroom (for example on a blackboard) where paper and pencil evaluations occurred.

On balance, most of the current items address issues that students are well positioned to address: they ask about instructor behavior that can easily be observed and elements of the learning experience that students are well placed to assess.

Structure of Items

It is possible to look at the relation between the 19 Likert items that form the bulk of the iEval questionnaire. To the extent that these items vary independently- for example with some tending to be high when others tend to be low- each item may provide unique information. On the other hand, to the extent that the items all vary together- for example all tending to be high or low- each additional item may add little information.

A common way to understand the relationships among a group of items on a survey is factor analysis. A factor analysis was conducted with using the set of all iEval responses collected for undergraduate courses in the spring of 2014. Full results of this factor analysis are provided in Appendix Two.

These analyses suggest that current items can be grouped into three general areas: instructor behaviors, course materials, and student behaviors. The group of items related to instructor behavior explains most of the overall variation in evaluation scores. The practical implication is that those items that relate only to a single factor could likely be eliminated without greatly impacting the total amount of information provided by iEval.

Recommendations:

These analyses suggest that it may be wise to:

- 1) Consider expanding the number of choices on Likert scale from five to seven.
- 2) Consider rewording the current item about supplemental materials to exclude specifically electronic resources and, at the same time, adding one or two items specifically to address the use of this kind of technology in the classroom.
- 3) Consider removing most or all of the items related to student behaviors and attitudes. (Assuming that data from these items will not be disaggregated nor used in a more refined way.)
- 4) Consider removing some items about instructor behavior and course materials.
- 5) Consider removing the five items where the text of the question is not provided.

Taken together recommendations 3-5 would result in a survey that is shorter than at present. It may be wise to consider options that might make this streamlined survey easier to administer, for example, by taking steps to make the online interface mobile-friendly.

Appendix One: Sample Questionnaire (formatting approximates student's online view)

SECTION I: A. Student Profile

1: I had a strong desire to take this course				
Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
2: I attended class regularly				
Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
3: I put considerable effort into this course				
Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
4: I gained a good understanding of the course content				
Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
5: I normally spent at least two hours preparing for each hour of class				
Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree

B. Instructor Appraisal

6: Instructor was prepared and organized

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

7: Instructor used class time effectively

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

8: Instructor was clear and understandable

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

9: Instructor exhibited enthusiasm for subject and teaching

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

10: Instructor respected students; sensitive to and concerned with their progress

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

11: Instructor was available and helpful

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

12: Instructor was fair in evaluating students

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

13: Instructor was effective as a teacher overall

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

C. Course Appraisal

14: The syllabus clearly explained the structure of the courses

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

15: The examinations reflected the materials covered during the course

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

16: The required readings contributed to my learning

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

17: The assignments contributed to my learning

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

18: Supplementary materials (e.g. films, slides, videos, demonstrations, guest lectures, iLearn, web pages, etc) were informative

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

19: The course overall as a learning experience was excellent

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

D. Department-Specific Issues

This section should only be completed if your instructor has provided you with course specific questions for this evaluation. The instructor may provide the questions in class, on iLearn, or in an email. If you did not receive any questions from the instructor please skip to Section II.

20: Q1

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

21: Q2

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

22: Q3

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

23: Q4

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

24: Q5

Strongly Disagree	Disagree	Neutral / Not Sure	Agree	Strongly Agree
--------------------------	-----------------	---------------------------	--------------	-----------------------

SECTION II:

25: Please comment on how the instructor's teaching helped your learning of the material in this course. Please give serious thought to your comments. Your comments will be studied by the professor after the grade and performance evaluation of your work have been submitted and may be used in changing future offerings of the course. In addition, these comments are placed in the instructor's file and may be used for purposes of evaluating the instructor's teaching. The information collected will remain anonymous.

Current Count: 0; Limit: 4000



Submit

Appendix Two: Factor Analysis

This section presents the results of a factor analysis using the set of all iEval responses collected for undergraduate courses in the spring of 2014. This includes 38,394 evaluations (submitted by 12,963 students because students typically take three or four courses).

Factor analysis assumes that the variability observed in a group of survey items is caused by two sources. One is an underlying but not directly observed factor that influences individuals' responses across a number of survey items. For example, the iEval instrument aims to measure effective teaching and to the extent that this is present student responses to several items should vary together. The other source of variation is an error term. To extend the example, bias would reduce the ability of the same group of items to measure effective teaching.

The figure on the next page presents the results of a second confirmatory factor analysis with three factors. Specific items are shown in the boxes along the right side with larger circles at the left showing the underlying factors; the small circles at the far right indicate error terms. The values along the straight arrows connecting the factors to the items indicate the proportion of variance in a given item that can be explained by the factor. The first factor explains much of the variance (50% or more) for each of the items that ask about instructor behaviors and attitudes as well as the item for the overall course experience. It also explains some of the variance (35- 45%) for the items about the syllabus and exams. The second factor explains much of the variance on the item that asks about the quality and clarity of course materials (very roughly 45- 85%). While this factor explains a fair amount of the variance in items about the syllabus note the variance here is also explained by the factor related to instructor behavior. The factor for course materials also explains 39% of the variance in the item asking about the course overall. The third factor explains variation in the group of items asking students about how much time they put into the course and how much they learned.

