

Response Bias in Paper vs PICA Course Evaluations
Mark Troy
December 2014

The observation of some faculty doing evaluations in both paper and in PICA is that ratings submitted in PICA tend to be lower than ratings submitted on paper evaluation forms. This paper examines the truth of that observation. The validity of the most common explanation for the difference is tested and an alternative explanation is proposed.

Alternative explanations for lower mean ratings in PICA and higher mean ratings on paper.

1.) The non-response bias explanation. This hypothesis is based on a common observation that response rates to online systems are lower than response rates to paper systems. It assumes that students who do not submit evaluations are different than those who do submit. The expectation is either that students who are motivated to submit evaluations are more likely to have a negative attitude towards the course or instructor or that students who stay away have a more positive attitude. Sometimes the explanation is that the most highly motivated students are both those with the strongest negative attitudes and those with the strongest positive attitudes.

Regardless of the motivation that is posited, the hypothesis is that the lower response rate produces a shift in the distribution of scores, such that more low scores than high scores are submitted.

2.) The response-set bias explanation. This hypothesis is based on an observation that when survey questions are presented in a matrix format, some respondents will adopt a response set known as "straight lining." Matrix formats are common on bubble forms. See the attached example of a bubble form used for evaluations. Questions are arrayed one below the other and the response options are presented in columns. In "straight lining" respondents simply mark the same response to each item giving the effect of a vertical line of bubbles.

This hypothesis also assumes that "straight lining" is the option chosen when respondents are not motivated sufficiently to consider each question separately on its own terms and that students tend to view their instructors in an overall positive light, also known as a "halo effect." The result is that "straight line" responses tend to be high.

It should be noted that the appearance of a straight line response, in and of itself, is not an indication that students are insufficiently motivated to give thoughtful responses to each item. A student, after careful consideration of each item, might assign the same rating to each, which would be visually and statistically identical to a straight line. The hypothesis is that the matrix format would yield more straight line responses than a non-matrix format.

These two hypotheses were tested with student ratings data from one college in 2013.

As was stated above, the paper evaluations have a matrix format, an example of which is given below. The PICA form is formatted differently. An example of part of the PICA form is presented below. The essential differences between the paper format and PICA are:

1. Questions on paper are printed on blank lines, one above the other, with the response options on the right side of the page. Questions in PICA are presented separately with the response options below the question.
2. The response options on paper are arrayed horizontally to form side-by-side vertical columns. The response options in PICA are arrayed vertically below the question.
3. A comment box is presented below each set of response options in PICA. No comment boxes are present on paper.

The effects of the differences between paper and PICA are that students require more time to read the items in PICA and are forced to view each one before making a response.

Data

Student ratings data were obtained from one large college in fall 2013. The enrollment in all courses that were appraised totaled 28,220 students. Evaluations were received from 21,006 or 74% of the total. This particular college uses paper evaluations for about 60% of courses and PICA for the remaining 40%. The average response rate for the paper evaluations was 74.9% and for PICA, 47.5%.

Analysis

All courses in this college are evaluated with the same eight questions. Each item is scored on a 5-point Likert scale. The unit of analysis in this study was the student mean score. That is, the average score given by a student over all eight items. The means of the student means are given in Table 1. As can be seen in Table 1., the mean ratings given on paper were almost half of a point higher than the mean ratings in PICA. On a five-point scale, this is a significant difference.

Table 1. Mean Ratings, Fall 2013

Overall mean	4.08
Paper mean	4.23
PICA mean	3.80

If non-response bias is a factor in the ratings such that only students with negative views are more likely to submit evaluations online, then we would expect to see a positive correlation between the response rate for a course and the mean ratings given by students in the course. The correlations are reported in Table 2. The overall correlation is small. The correlations for each mode of evaluation are even smaller and approach very close to zero.

Table 2. Pearson Product-Moment Correlations Between Mean Ratings and Course Response Rate

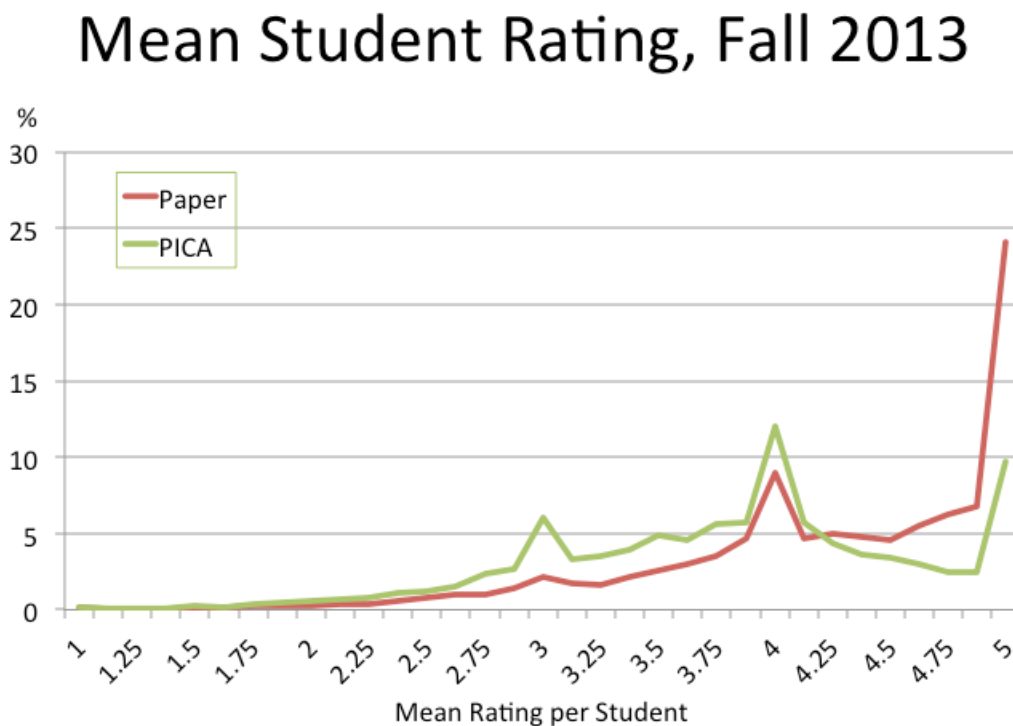
Overall correlation	0.108
Paper correlation	0.048
PICA correlation	0.054

The correlations could be near zero if the distributions were bimodal. That is, if “Only those who like the instructor and those who really hate the instructor” submitted evaluations. This might account for the low PICA correlation, but how does it account for the low paper correlation?

To determine if the distributions are bimodal, the percentages of students giving each mean rating were plotted for paper and PICA separately.

The results are shown below in Figure 1.

Figure 1.



As can be clearly seen from the chart nearly 25% of evaluations submitted on paper have a mean of 5.0 compared to about less than 10% of evaluations submitted in PICA. Clearly, neither paper nor PICA distributions are bimodal. Both are negatively skewed, though the paper distribution is more so. Responses at the low end of the distribution are nearly absent for both formats. Also noteworthy is the large discrepancy between the percent of 5's and the percent of 4's given on paper. It is actually larger than the difference in percentage of A's between paper and PICA. In contrast, the percentage of 4's on PICA is larger than the percentage of 5's in PICA. In general, there is an overall shift of responses towards the more moderate ratings in the PICA evaluations, suggesting that students are using more of the scale and, we hope, giving more thought to the evaluations online than on paper.

The analysis did reveal more instances of straight-lining on paper than on PICA. All scores of 5.0 are straight lines as are all scores of 1.0. Not all scores of 2.0, 3.0, and 4.0 result from straight-lining. We defined a straight line is a set of responses such that the minimum, maximum, and means are equal. The percentage of straight lines for each score are given in Table 3. These percentages parallel those shown in Figure 1.

Table 3. Straight lines by evaluation format.

Straight line	Paper	PICA
A	24.12%	9.73%
B	5.69%	8.30%
C	0.96%	3.94%
D	0.04%	0.14%
E	0.18%	0.18%
Total	30.99%	22.3%

It could be argued that the above differences could occur if only students who dislike the instructor completed the evaluation in PICA. In that case we might see a shift to lower means. In order to test that, the analysis was run again but limited to courses that had a 50% or higher response rate. The reasoning being that a non-response bias would not be a factor in courses with high response rates. And we would expect the PICA results to look more like the paper results. Similarly, if we limit the analysis to courses that had a response rate below 50%, we might see even greater discrepancies. The results are presented in Figures 2 and 3 below. The mean response rates were: paper = 79.1% and PICA = 74.1% in the 50 and above group; paper = 29.2% and PICA = 34.3% in the below 50 group. As can be seen, the two charts are nearly identical with Figure 1. We can conclude that the difference between paper and PICA response patterns is not due to non-response bias, but to the response set created by the matrix format on the paper forms.

Figure 2.

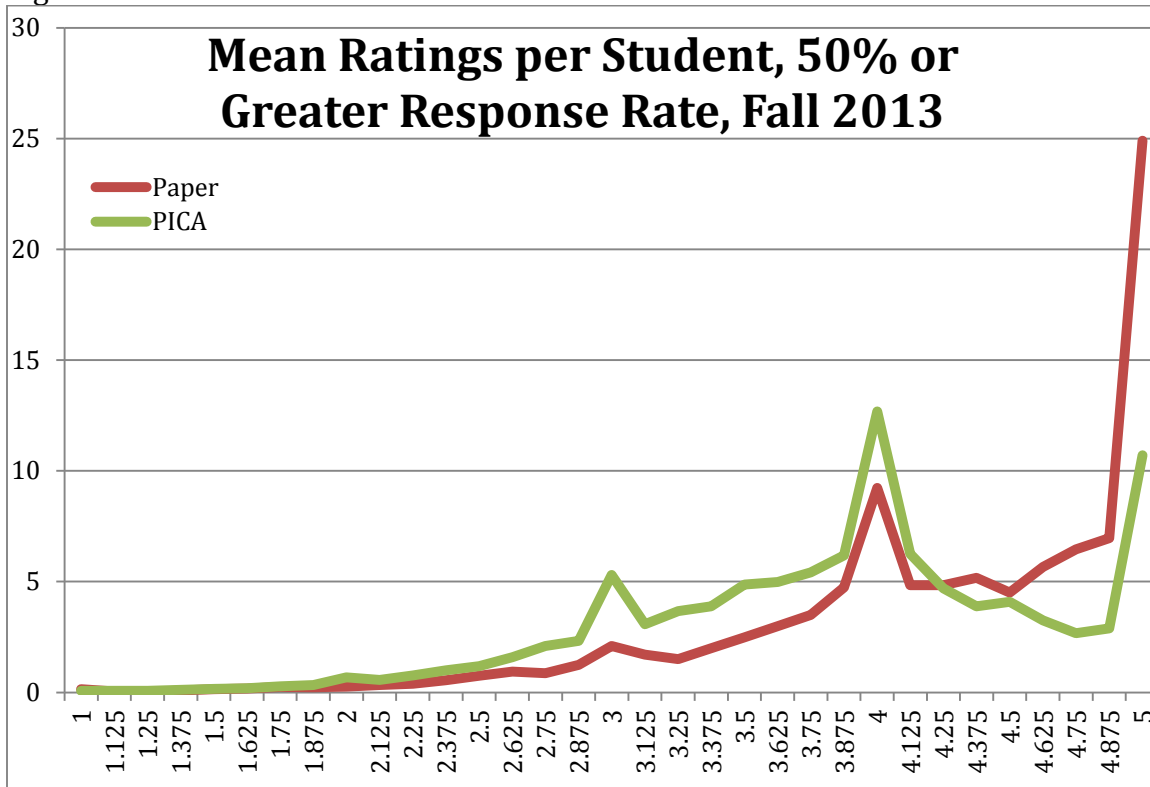
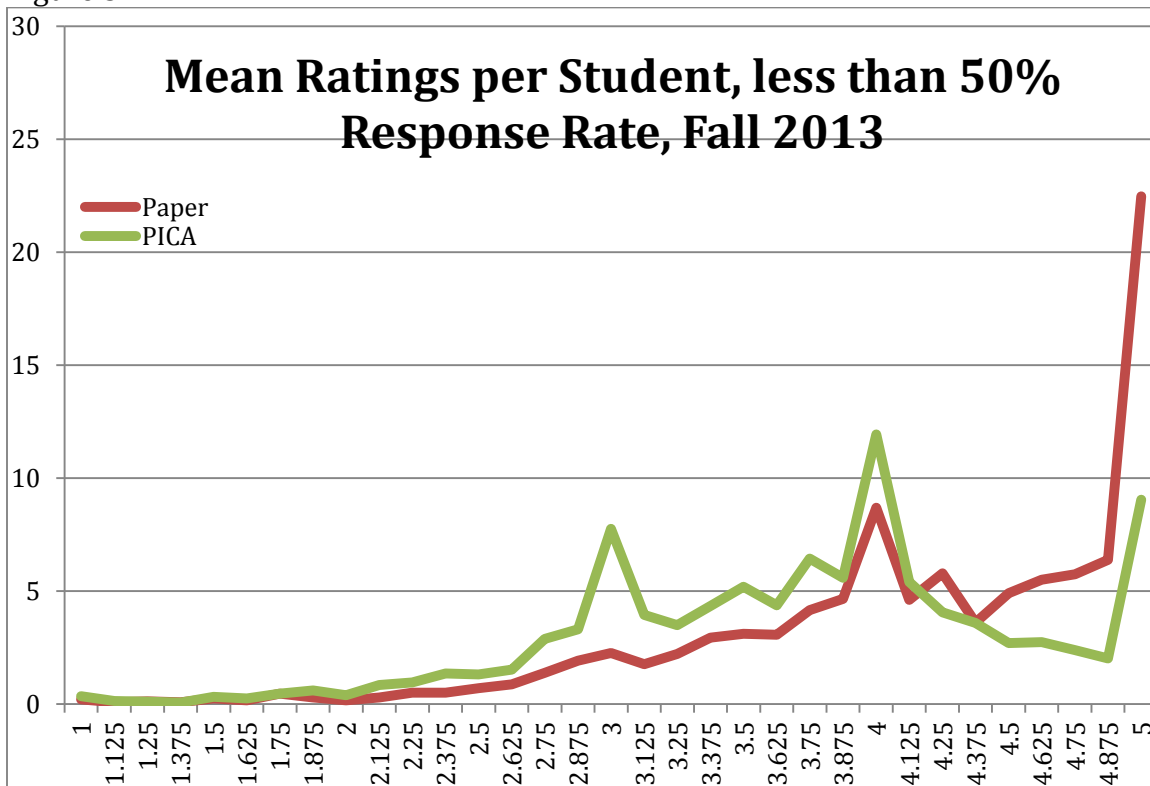


Figure 3.



Conclusions

This analysis began with the hypothesis that mean scores on paper evaluations tend to be higher than mean scores on PICA evaluations. The difference in mean scores was confirmed for this particular set of data.

Two hypotheses were offered to explain the difference. The first hypothesized a non-response bias, which takes the form of an assertion that only students who hate the instructor will do the evaluation, or only students who really like the instructor or really hate the instructor will do the evaluation. Either way, this hypothesis proposes a shift of response distributions such that lower scores are associated with lower response rates.

The results clearly discredit this hypothesis. A significant correlation would be required to support the hypotheses, but the observed correlation was not only not significant, it was near zero, meaning there is no relationship between ratings and response rate. If the hypothesis were true, we would expect PICA scores in courses with high response rates, i.e., those for which a non-response bias doesn't exist, to be similar to scores obtained from paper evaluations, and we would expect PICA scores in courses with high response rates to be different from PICA scores in courses with low response rates. Neither of those outcomes were found. PICA scores differed significantly from paper scores in both the high response group and the low response group and the distributions were nearly identical in both groups.

The other hypothesis is that the format of the items and responses on paper tend to encourage the development of a response bias popularly known as "straight-lining." Straight-lining is recognized as a result of poor questionnaire design in which items and the responses are arranged on a matrix or grid format. Even though a grid is a set of independent items, respondents often respond to a grid as a single item and provide respond with a single response. This is especially true if their motivation to take the survey is low to begin with.

Other biases come into play as well. One of those biases is known as the acquiescence response bias. This bias results from the desire of many people who take surveys to want to be seen as likeable and agreeable. Consequently, they are more likely to choose the "Strongly agree" or "Agree" options than any other. Students are also susceptible to another bias in student ratings know as the "halo effect." Basically students like and admire their teachers and are inclined to rate them highly or at least give the benefit of the doubt. When it comes student ratings on paper, the straight-lining, acquiescence bias, and halo effect bias, combine to produce an excessive number of "A" responses.

You could say that paper evaluations are given to a captive population. Students are in class and have to complete it before leaving. Many of those students don't care about doing the evaluations and are, therefore, more likely to do a straight-line response just to get the

evaluation over with. For them completing the evaluation is a ticket to leave. Students doing online evaluations have to make an effort to go on line and do so voluntarily. They have more interest in doing the evaluations and are likely to pay more attention to the questions. The difference in response rate between paper and PICA was about 24% and the difference between the straight A's in paper and PICA is 14.39%. Some students didn't do the online evaluations because they forgot ("I forgot" is the most common excuse for not doing online evaluations). If the difference between the straight A's is an indicator of the students who don't care, then the overall difference in response rate can be attributed to a combination of students who don't care and students who forgot.

Two conclusions can be drawn from the data: First, the non-response bias, that only those who really like you and those who really hate you will complete the evaluation online, doesn't exist.; and, Second, the design of PICA successfully overcomes a design flaw in the paper design which induces biased responding that inflates scores.

The design of PICA incorporates features intended to force students to be slower and more deliberate in rating their courses and instructors. Although fewer responses are obtained in PICA than on paper, the available evidence indicates that they result from more thought than responses given on paper, and are, therefore, more likely to have higher reliability and validity.

Mark Troy
metroy@tamu.edu

References

<http://survey.cvent.com/blog/customer-insights-2/ozzy-osbournes-advice-for-questionnaire-design>

<https://www.surveymonkey.com/blog/en/blog/2014/09/19/2-tips-for-writing-agree-disagree-survey-questions/>

<https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-GRBN-draft-Online-Sample-Quality-Guideline-April-2014.pdf>

<http://lovestats.wordpress.com/2009/03/25/survey-design-tip-3-do-you-encourage-straightlining/>

Paper Evaluation Form

TEXAS A&M UNIVERSITY
 MEASUREMENT AND RESEARCH SERVICES

INSTRUCTOR AND COURSE APPRAISAL

PLEASE RESPOND TO ANY ITEMS IN THIS SECTION BY MARKING THE APPROPRIATE SYMBOLS.
 USE NO. 2 PENCIL. ONLY FIVE CHANNELS ON CORNER OF PENS SHOULD BE USED.

ARE YOU?			EXPECTED COURSE IN			FOR INSTRUCTOR USE ONLY					
MAJOR YOUR COLLEGE:	<input type="checkbox"/>	YES	FRESH	<input type="checkbox"/>	SO	1	2	3	4	5	SALARY RESPONSE
AGRICULTURE	<input type="checkbox"/>	NO	JUNIOR	<input type="checkbox"/>	2	3	4	5	6	7	
ARCHITECTURE	<input type="checkbox"/>		SENIOR	<input type="checkbox"/>	3	4	5	6	7	8	
BUSINESS ADM.	<input type="checkbox"/>		GRAD	<input type="checkbox"/>	4	5	6	7	8	9	
EDUCATION	<input type="checkbox"/>		OTHER	<input type="checkbox"/>	5	6	7	8	9	10	
ENGINEERING	<input type="checkbox"/>		PASS/FAIL	<input type="checkbox"/>	6	7	8	9	10	11	
GENERAL STUDIES	<input type="checkbox"/>		ADULT	<input type="checkbox"/>	7	8	9	10	11	12	
GEOSCIENCES	<input type="checkbox"/>		OTHER	<input type="checkbox"/>	8	9	10	11	12	13	
HUMANITIES	<input type="checkbox"/>		IS THIS COURSE REQUIRED?	<input type="checkbox"/>	9	10	11	12	13	14	
LIBERAL ARTS	<input type="checkbox"/>		YES	<input type="checkbox"/>	10	11	12	13	14	15	
MEDICINE	<input type="checkbox"/>		NO	<input type="checkbox"/>	11	12	13	14	15	16	
SCIENCE	<input type="checkbox"/>		SEX:	<input type="checkbox"/>	12	13	14	15	16	17	
VET. MEDICINE	<input type="checkbox"/>		MALE	<input type="checkbox"/>	13	14	15	16	17	18	
OTHER	<input type="checkbox"/>		FEMALE	<input type="checkbox"/>	14	15	16	17	18	19	
PLEASE READ EACH STATEMENT CAREFULLY, THEN SELECT ONE OF THESE FIVE ALTERNATIVES.											

* * PLEASE WRITE COMMENTS ON REVERSE SIDE OF THIS FORM * *

Person 8888711-1 45443503140

PICA Appraisal Form

Instructor and Course Appraisal

Note: The system will accept a maximum of 1500 characters for each comment in this section of the appraisal.

Please read each statement carefully, and then select the most appropriate answer from one of the available five choices: ➊

1. The instructor discussed recent developments in the field.

Strongly Agree
 Agree
 Undecided
 Disagree
 Strongly Disagree

Please elaborate with examples or additional comments:

2. The instructor showed concern for the quality of his/her teaching.

Strongly Agree
 Agree
 Undecided
 Disagree
 Strongly Disagree

Please elaborate with examples or additional comments: