



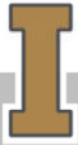
The purpose of this presentation is to present the findings from the initial comparison study of the new instructor and course items as part of a desired revision of the Student Feedback form.

## PURPOSE: INITIAL STUDY

There was a request to revise the items and form used to obtain student feedback on instruction and course. Concerns about the comparability of the old data and items to the new was of importance for longitudinal uses such as promotion and tenure. A process to examine this and other concerns about the form was developed. The experimental form contained the two new items on instructor and course, along with the previous teacher and course items. Previous work with the old form had used additional items as a validation reference and items used here were selected to serve the same purpose. A summary of this previous study is available online at:

<http://www.uidaho.edu/provost/iea/accreditation-evaluation/student-evaluation>.

The revised form and items were administered campus-wide in the Fall of 2016 and the Spring of 2017. This report makes use of these data.



University of Idaho

General history of the process that led to this study.

## GENERAL REVIEW

The SET form is not intended to be used in isolation, but is required as part of a process to examine both teaching and course structure.

There is an expectation that each College has an established teaching and course review process, of which this is to be one part.

- IEA recommends that this process include other evaluation procedures for teaching and course assessment to supplement and provide context for the results of this tool.

For example, the tool is part of the process which may be used at the university level in the context of DFWI reports (included in these are course GPA and % A grades) and other academic measures. The tool is not intended to be reviewed in isolation.



This contextual use assists in providing a checks-and-balances process to ensure rigor in coursework as well as examination of teacher and course elements.

University of Idaho

Some general reminders about the use of Student Feedback forms.

Though not shown here, the distributions for these instructor and course items were examined. They are quite similar and consistent with what was found in researching the old items. The distributions for both the old and new items suggest these items are better able to discriminate lower scores from average or above, than at discriminating higher scores from average.

## REVIEW: PERSONAL VALUES

There are few areas in higher education that have the emotional loading that a student evaluation of teaching carries. Nearly every faculty member believes himself/herself to be an excellent instructor with well-run and interesting courses. There are few “average” self-ratings here.

- Practically every instructor can locate a student in a course to provide support for them.
- Those wishing to find negative support can generally find a student to support that view as well.
- There has been enough research in this area (generally very poorly designed with few controls for extraneous variables) that one can find support for nearly any position. Much like student support, one can also find the support of an opposing view in the literature as well.



University of Idaho

Another set of reminders about these examinations.

## REVIEW: SELECTED STUDIES

### Selected studies of interest

- Al-Busaidi, Aldhafri, S. & Buyukyavuz, O. (2016). Effective university Instructors as perceived by Turkish and Omani university students. *Sage Open*, July-Sept., 1-8.
- Marsh, H. & Roche, L. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity or innocent bystanders. *Journal of Educational Psychology*, 92(1), 202-228.
- Overall, J. & Marsh, H. (1980). Students' evaluations of instruction: A longitudinal study of their stability. *Journal of Educational Psychology*, 72(3), 321-325.
- Pietrzak, D. Duncan, K. & Korcuska, J. (2008). Counseling students' decision making regarding faculty teaching effectiveness: A Conjoint Analysis. *Counselor Education and Supervision*, 48(2), 114-132.



University of Idaho

Some foundational research that may be of interest.

## REVIEW SELECTED STUDIES CONTINUED

### Selected studies of interest continued

- Remedios, R. & Lieberman, D. (2008) I liked your course because you taught me well: The influence of grades, workload, expectations and goals on students evaluations of teaching. *British Educational Research Journal*, 34(1) 91-1125.
- Maurer, T. (2006). Cognitive Dissonance or Revenge? Student Grades and Course Evaluations. *Teaching of Psychology*, 33(3), 176-179.
- Sporeen, P. & Mortelmans, D. (2006). Teacher professionalism and student evaluation of teaching: will better teachers receive higher ratings and will better students give higher ratings? *Educational Studies*, 32(4) 201-214.
- Addison, W., Best, J. & Warrington, J. (2006). Students' perceptions of course difficulty and their ratings of the instructor. *College Student Journal*, 40(2).

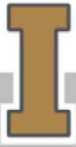


University of Idaho

More research.

## REVIEW: SUMMARY

Summary – While it is often argued that grades and other elements should have no relationship to instructor evaluations this is not theoretically expected. Theoretically it is generally believed there should be some relationship. For example, theoretically there should be some relationship between grades and ratings. Instructors that are experienced by students as assisting with student learning should have higher performing students. This is not really a new development. Also note that no single assessment measures one thing. The question really is, “Does the effect of these other elements negate the utility of these types of tools?” These need to be controlled in studies. However, the current data seem to suggest these forms pick up largely a “classroom environment” factor.



University of Idaho

These type of tools appear to function very similarly regardless of publisher. These tools seem to pick up how a student experiences the instructional atmosphere of the classroom. This appears largely an instructor-influenced and instructor-controlled element. The student responses are addressing the question “Did this instructor assist me in my learning?”

There are multiple instructional delivery systems and it is hoped that instructors are a value-added component of the process. As such, there should be some relationship to grades, but that should have a relatively small influence.

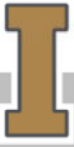
We see other aspects that are expected to have some relationship to these scores. Smaller classes should provide a greater opportunity for the instructor to assist the student. As such we expect there to a positive relationship to class size.

Though not shown here, the distributions for the instructor and course items were examined. These distributions are quite similar to what was found with the research on the old items and form. We will examine these more closely in later slides.

Lastly, the distributions of both course and instructor items (both old and new) suggest these items are better at discriminating lower scores from average or above, than at discriminating higher scores from average.

## GENERAL METHOD

This is an initial examination intended to serve as a means to compare the new teacher and course items to the old teacher and course items (alternate form reliability). In addition to comparison with each other (alternate form reliability), the relationship of the key's "optional" items were selected based on previous research using the old form. These are intended to serve as a means to investigate the evidence of convergent and discriminant validity. As a means of general methodology a Multitrait-Multimethod model guides the process (<https://www.socialresearchmethods.net/kb/mtmmmat.php>). The test form was administered in the Fall of 2016 and Spring of 2017 to a total of 33,186 students, representing 4,718 courses at both the graduate and undergraduate levels. With sample sizes this large, small correlations are often statically significant ( $r = 0.04$  or larger). The practical significance needs to be accounted for here as well.



University of Idaho

The general philosophy that guides this research process is drawn from the nomological network (<https://www.socialresearchmethods.net/kb/nomonet.php>). This is not a new concept or approach. It has been enacted in various ways including several variations of Multitrait-Multimethod modeling (<https://www.socialresearchmethods.net/kb/mtmmmat.php>). This study approaches this process from one such model. However, the guiding principles are the same.

- 1.) Do the elements have relationships with things they are expected to have relationships?
- 2.) Do the elements not have relationships with things that they should not have relationships?
- 3.) Do the pattern of relationships in terms of direction and magnitude make theoretical sense?
- 4.) Are there unexpected relationships (size and direction) with other things that raise questions about the utility of the tool for its' intended use?

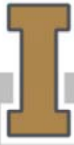
As medications rarely have a single action on the body, assessments rarely tap a single construct. For example, aspirin thins the blood **and** reduces the sensation of pain. The question is, does the medication's prescribed use have sufficient merit as to make it a viable application? So too with assessment, the question is, does the tool sufficiently measure what is intended without too much interference and/or noise? The various Multitrait-Multimethod models provide general guidelines for use of such an examination.



## EFFECT SIZE (COHEN) SUGGESTED FOR SOCIAL SCIENCES

<u>Pearson</u>	<u>R<sup>2</sup></u>
Small +/- 0.2	Small 0.04
Medium +/- 0.5	Medium 0.25
Large +/- 0.8	Large 0.64

***Note that in this study, essentially any correlation (r) of 0.04 or larger is statistically significant. This amounts to 0.2% shared variance leaving 99.8% unaccounted for which has no, or little, practical utility.***



From (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3444174>)

University of Idaho

The role of statistical significance has been widely misunderstood by many. It is based on a combination of sample size and the specific statistical technique's power in finding differences that are likely to be greater than chance. However, this has been widely misused over the years and has led the majority of research journals in the social sciences to stress the importance of effect size, or practical significance. While it is certainly important to examine statistical significance, it is not sufficient when considering the utility of the results. Effect size is an attempt to consider the practical implications of the research results. While it can be misused as well, it allows for a more balanced consideration of the meaning of the results. One can say that while statistical significance is necessary, it is not a sufficient condition for examining the practicality of any results.

# ALTERNATE FORM RELIABILITY: COURSE LEVEL

Instructor and Course Items



University of Idaho

The scores for items such as these are reported based on class averages. As UI only reports such scores where there are 5 or more respondents, this will be the focus here. A key step is to determine if the 2 items (old and new) are sufficiently related to be considered alternate forms of one another.

## ALTERNATE FORM RELIABILITY (RESPONSES OF 5 OR MORE)

(N=2353)

<b>Alternate Form Reliability</b> The generally accepted guideline: .90 > excellent .80 -.89 good .70 -.79 adequate Below .70 limited applicability <a href="http://www.hr-guide.com/data/G362.htm">http://www.hr-guide.com/data/G362.htm</a>	16. (Old) Course	17. (Old) Instructor	4. (New) Course	5. (New) Instructor
16. (Old) Overall, how would you rate the quality of this course?	1.00			
17. (Old) Overall, how would you rate the instructor's performance in teaching this course?	0.92	1.00		
4. (New) Overall, the content and organization of this course contributed to your understanding of this subject	<b>0.91</b>	0.88	1.00	
5. (New) Overall, the instructor's delivery and efforts contributed to your understanding of the course material.	0.90	<b>0.93</b>	0.93	1.00



Reliability coefficients are direct shared variance measures. Example:  
0.92= 92% shared variance.

University of Idaho

These are the results from examining mean scores from courses where there are 5 or more responses. Here we see that the new items have a very strong relationship to the similar content old items. However, there is little distinction between the course and instructor scores. These relationships provide strong support for the items addressing the same construct.

These data would not support considering the course and instructor items as measuring different constructs in an applied setting. They are too highly related to one another to support discrimination from one another. The total N here is 2353.

## ALTERNATE FORM RELIABILITY (ALL COURSES)

<b>Alternate Form Reliability</b> The generally accepted guideline (bold text): .90 > excellent .80 -.89 good .70 -.79 adequate Below .70 limited applicability <a href="http://www.hr-guide.com/data/G362.htm">http://www.hr-guide.com/data/G362.htm</a>	16. (Old) Course	17. (Old) Instructor	4. (New) Course	5. (New) Instructor
16. (Old) Overall, how would you rate the quality of this course?	1.00	4684	4690	4690
17. (Old) Overall, how would you rate the instructor's performance in teaching this course?	0.88	1.00	4684	4684
4. (New) Overall, the content and organization of this course contributed to your understanding of this subject	<b>0.84</b>	0.80	1.00	4694
5. (New) Overall, the instructor's delivery and efforts contributed to your understanding of the course material.	0.84	<b>0.87</b>	0.89	1.00



Reliability Coefficients below diagonal with total courses above the diagonal.  
Reliability coefficients are direct shared variance measures. Example: 0.88= 88% shared variance.

Here we see the same as in the previous slide. As expected when including courses with low number of responses, the relationships decline slightly. This provides support for the continued reporting of courses with 5 or more responses. Note in this case the reliability coefficients are presented below diagonal with total courses for each relationship are above the diagonal. The variation is due to some items being left blank by some students.

## STUDENT LEVEL CORRELATION

It is generally expected that the correlation at the student level on these items will be lower than at the course level. These correlations provide a means to estimate the degree of shared meaning across students and ratings. All student responses are included in this estimate. This supports that there is largely general agreement on understanding and ratings. Here we see greater agreement on the instructor than the course.

Student Consistency	16. (Old) Course	17. (Old) Instructor	4. (New) Course	5. (New) Instructor
16. (Old) Course	1.00	32,259	32,450	32,450
17. (Old) Instructor	0.85	1.00	32,324	32,324
4. (New) Course	0.78	0.76	1.00	32,654
5. (New) Instructor	0.78	0.83	0.86	1.00



Consistency Coefficients below diagonal total courses above the diagonal.  
Reliability coefficients are direct shared variance measures. Example: 0.88= 88% shared variance.

The variation in the meaning across students by item can be examined by comparing student level correlations to the course level correlations. This difference can be conceived of as an estimate of the degree of shared meaning of students for an item. It addresses the degree of shared meaning across students. It addressed the question "Do students generally read the item and interpret it much the same way?" As noted previously these data suggest there is little discrimination of the course from instructor score.

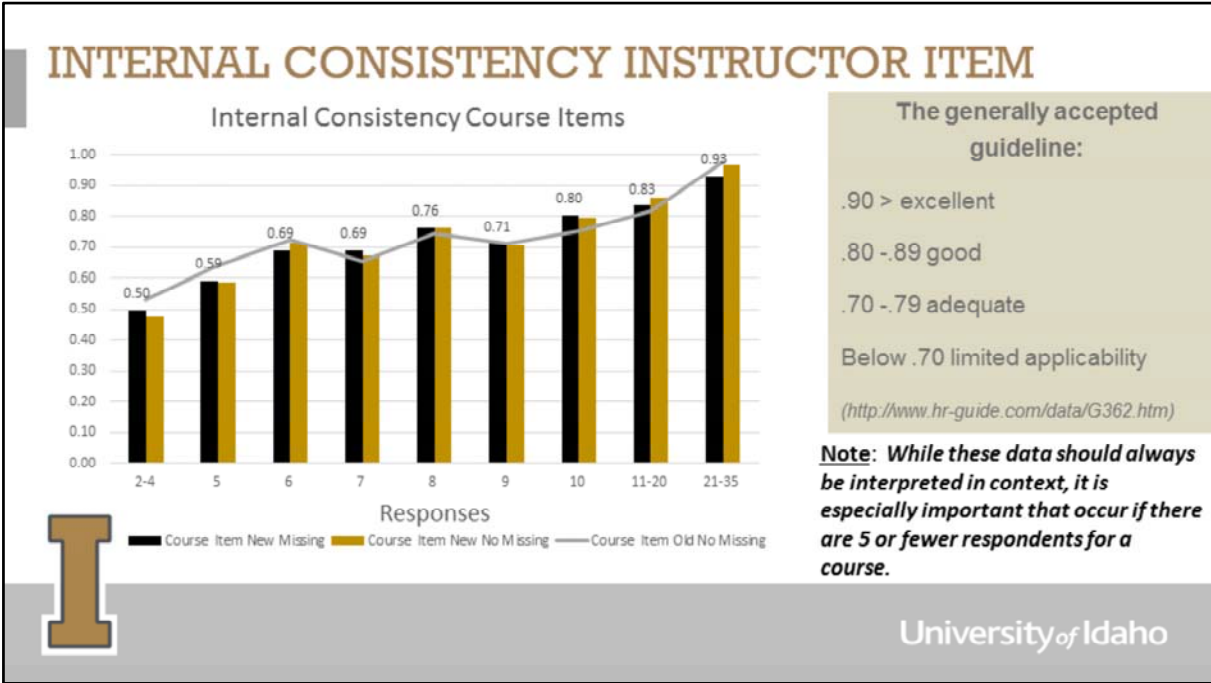
# INTERNAL CONSISTENCY (RELIABILITY): COURSE LEVEL

Instructor and Course Items



University of Idaho

The scores for items such as these are reported based on class averages. As UI only reports such scores where there are 5 or more respondents, this will be the focus here. A key step is to determine if the 2 items (old and new) are sufficiently related to be considered alternate forms of one another.



Internal consistency was calculated using Cronbach's Alpha (standardized). In these type of instruments, the number of students responding are treated as separate items that might appear on more traditional tests. Data like this is usually reported using groupings of respondents such as 2-4, 11-20, and 21-35 here. However, to better examine these in more detail, the number of students completing the form was used. The estimates were calculated 2 ways: 1.) only allowing examination if all respondents in a group responding to the question, and 2.) including all persons responding to the instrument even if someone skipped responding to the item. While the former is more traditional, the estimates including those who may have skipped questions tends to align with real life application.

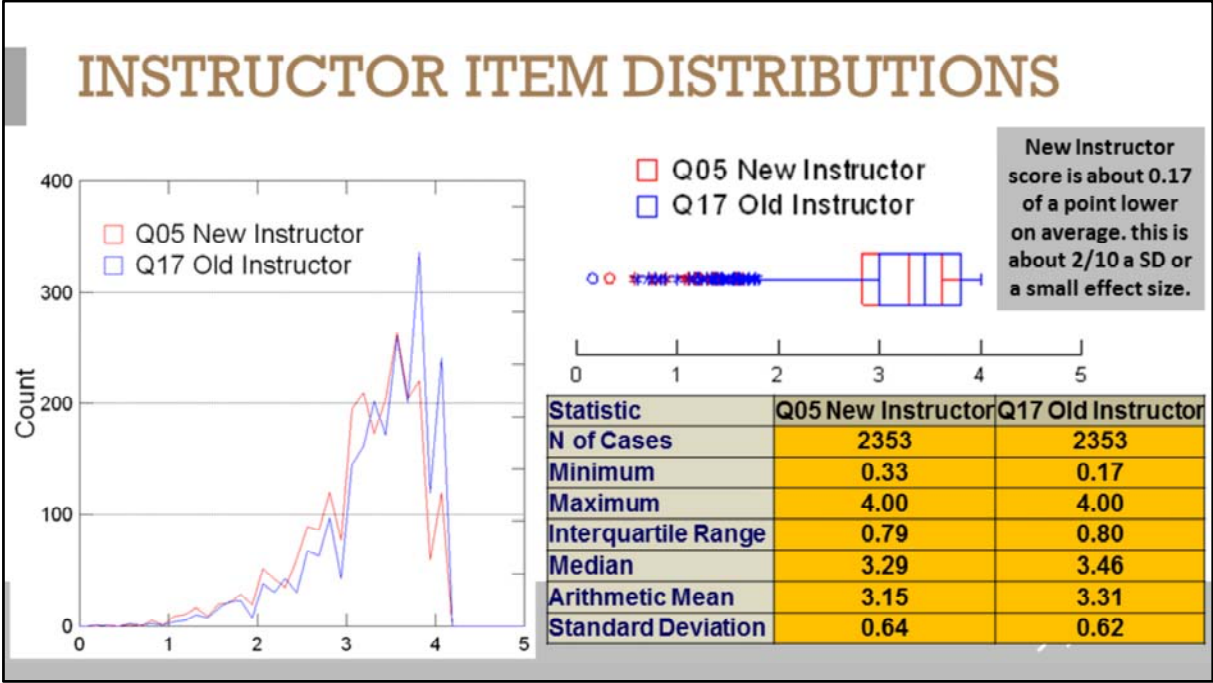
We see here that when there are 5 students who responded in a course, it has an internal consistency estimate in the 0.59 to 0.60 range. This is lower than one would like to see in general application if the scores are being interpreted independently of any other such feedback in a term or historical feedback. As such, when there are 5 or fewer responses it is especially important that the information be used in the context of other feedback in that term and/or historically.

## INSTRUCTOR AND COURSE ITEM DISTRIBUTIONS

- Comparisons of the mean & median. The standard deviation and interquartile range are considered. Additionally distributions of old and new items for instructor and course are examined.

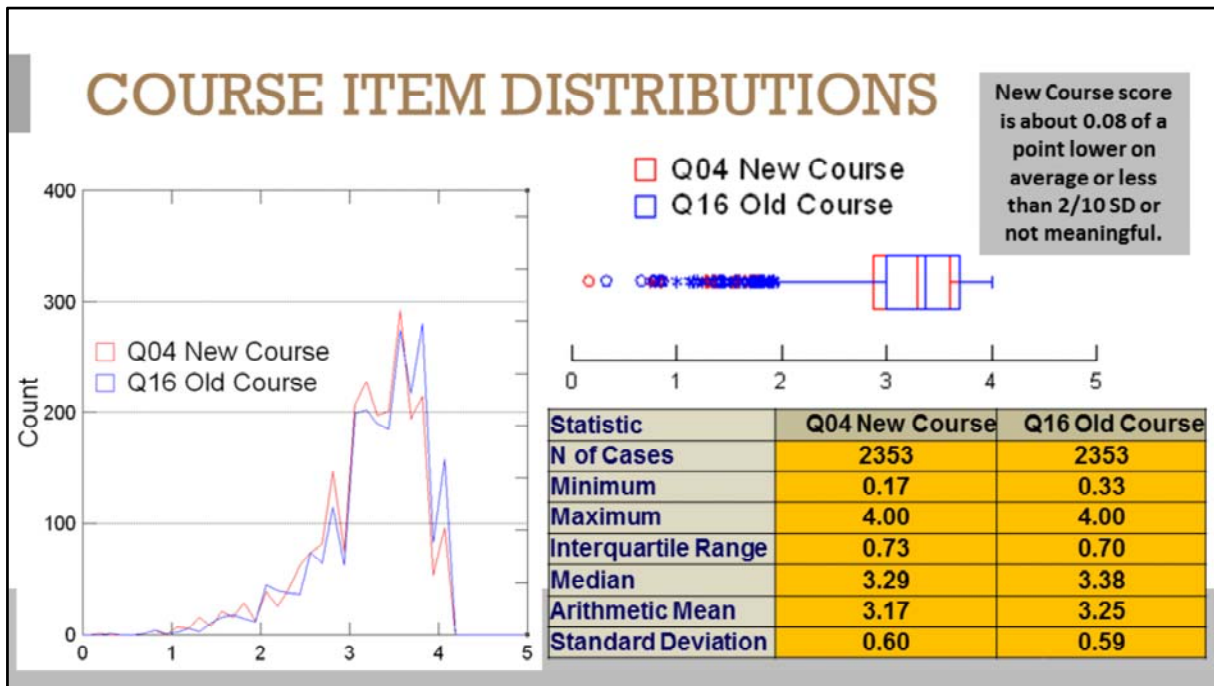






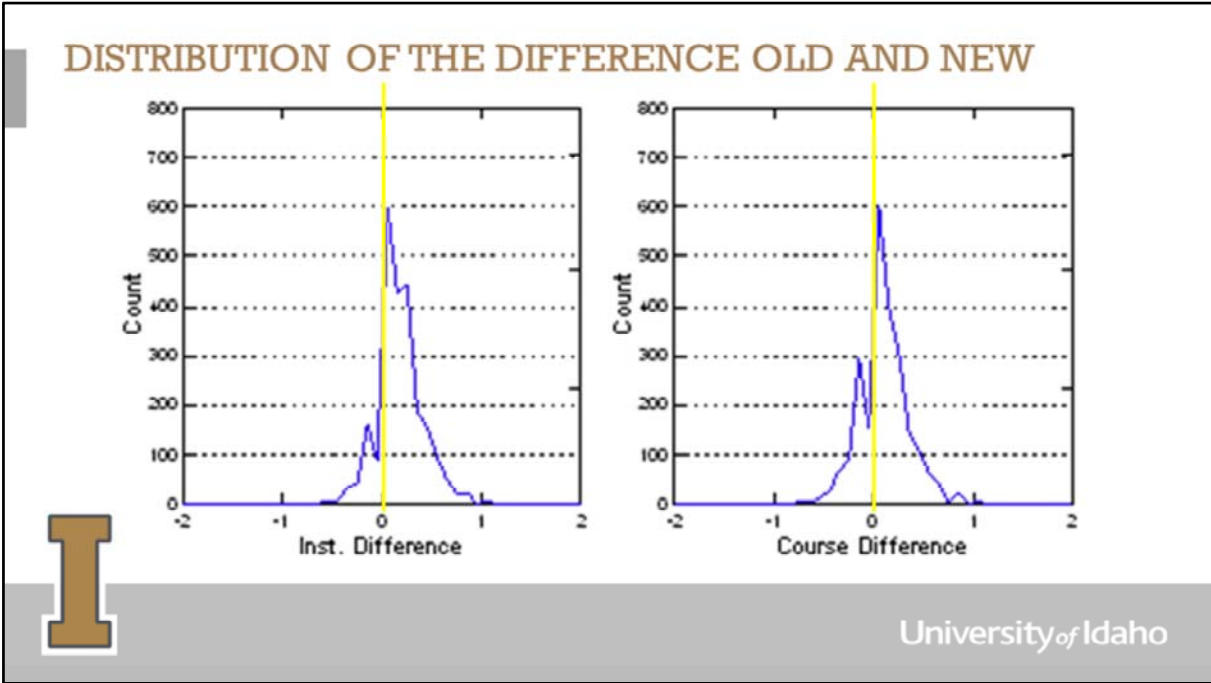
These data suggest that the new Instructor score on average is just under 0.20 points lower than the old score. This represents approximate 2/10ths of a standard deviation which is the low **bound** for a small effect size. The general shapes of the distribution are very similar as reflected in the similar standard deviations and interquartile ranges. The differences in the mean and median reflect the slightly lower average score on the new item.

\*\*DALE: is that supposed to be "bound" or "boundary"?



These data suggest that the new Instructor score on average is just about equal to the old score. The difference of 0.08 points difference between the new score from the old is well under  $2/10^{\text{th}}$  a standard deviation which is the low bound for a small effect size. The general shapes of the distribution are very similar which is reflected in the similar standard deviations and interquartile ranges. The differences in the mean and median reflect the very slight difference of the lower average score on the new item.

\*\*DALE: same here, bound or boundary??



These graphs display the differences in new and old course and instructor scores. These are relatively normally distributed but shifted off zero slightly indicating the scores for the new questions are slightly lower. The yellow line represent where zero falls.

This suggests that while the differences are small on average, we may want to consider correcting the scores by adding about 0.4 to the new instructor and course scores when looking at trends over time, including both old and new scores.

## INSTRUCTOR ITEMS: RELATIONSHIPS TO OTHER ITEMS AND DEMOGRAPHICS

- 5. (New) Overall, the instructor's delivery and efforts contributed to your understanding of the course material.
- 17. (Old) Overall, how would you rate the instructor's performance in teaching this course?



University of Idaho

Here we can examine the wording of the new item and older item in more detail. The newer item framed the student response with more detail. In doing so there is some blurring of the item content with course material in this case.

## ADDITIONAL ITEMS

01. How often did you attend class or an online learning environment?
02. How many hours per week (outside of class) did you do work for this course?
03. The instructor expressed clear expectations for learning outcomes in this course.
06. The instructor was helpful to me outside of class or in the online learning environment.
09. Clarity of instructor's explanations.
10. Likelihood you would recommend this instructor to others.
11. Instructor's ability to stimulate interest in the course topics
12. Presentation of course material by the instructor.
13. Course's value in gaining an understanding of the subject matter.
14. Appropriateness of level at which course material is covered.
15. Relevance of written assignments to course materials.



University of Idaho

Here we can see the list of additional items that are examined in relationship to the instructor and course questions. The relationships reported here are using Pearson correlations, save for item 6, which is a point-biserial correlation as it is binary (Yes, No with NA's excluded). Spearman correlations were examined in most cases as well, with the results being similar with the context that some were slightly lower and others slightly higher. There was overall little difference. Because the professional literature in this area typically reports Pearson and point-biserial, that is what is being reported here.

## INSTRUCTOR ITEM RELATIONSHIPS

### Correlations with Items to Old and New Instructor Scores: Top 5 (5+ Responses).

*These 6 items account for 91% of the variance of the New Instructor Item.*

	5. New Inst.	Rank New Inst.	17. Old Inst.	Rank Old Inst.
12. Presentation of course material by the instructor.	0.92	1.00	0.94	02
10. Likelihood you would recommend this instructor to others.	0.91	2.00	0.96	01
09. Clarity of instructor's explanations.	0.91	3.00	0.92	03
03. The instructor expressed clear expectations for learning outcomes in this course	0.90	4.00	0.87	07
13. Course's value in gaining an understanding of the subject matter.	0.89	5.00	0.90	05
11. Instructor's ability to stimulate interest in the course topics	0.89	6.00	0.92	04

*Effect Size: 0.2 small, 0.5 medium and 0.8 large*

University of Idaho

Convergent Examination: The relationships ranged from 0.92 to -0.04. From all of these items the top five relationships for the new and old instructor item are shown. These represent the upper limit of the relationship as the overlap with other items is not removed (not limited to unique variance). This supports that students in general read and responded to items independently (not just all items being marked the same). However, due to the limited number of items used here, further investigation would be needed in this area. However, given the correspondence of the new item with the old item it seems reasonable that we will find similar patterns for the old item. Previous research with the old item provided support for differentiated item responses by students.

Here we see that the more influential elements related to increased teacher scores were: 1) The presentation of the course material; 2) The instructor providing clarity and clear expectations for the course; 3) The courses value in gaining an understanding of the subject; 4) The likelihood that they would recommend the instructor to others (in the context of the elements described here); and, 5) The instructor's ability to stimulate interest. Here we see the impact of having elements of the course in the new item. The course element is included in the new items top 5 where the instructor's ability to stimulate interest moves down slightly. These differences are small but do suggest the course and instructor items are less differentiated now. However, they were not well differentiated with the old items.

Several multiple regression models were employed to look at the variance accounted for with the New Instructor item. The model with these 6 items accounted for 91% of the variance. Various combinations created by adding any other variables from the validity items and demographic variables increased the variance accounted for over this model. The full model with all of these variables stayed at 91%. The full model was able to account for an additional 0.04%, but due to the increased complexity actually saw a decline in the adjusted  $R^2$  of 0.01%. As the multicollinearity of these 6 items was high, the ability to discriminate the importance of each independent of the others in the model was not possible.

## INSTRUCTOR ITEM RELATIONSHIPS

Correlations with Items to Old and New Instructor Scores: Lowest (5+ responses)	5. New Inst.	Rank New Inst.	17. Old Inst.	Rank Old Inst.
06. The instructor was helpful to me outside of class or online learning environment.	0.58	9	0.62	09
01. How often did you attend class or an online learning environment?	0.29	10	0.26	10
02. How many hours per week (outside of class) did you do work for this course?	-0.04	11	-0.07	11

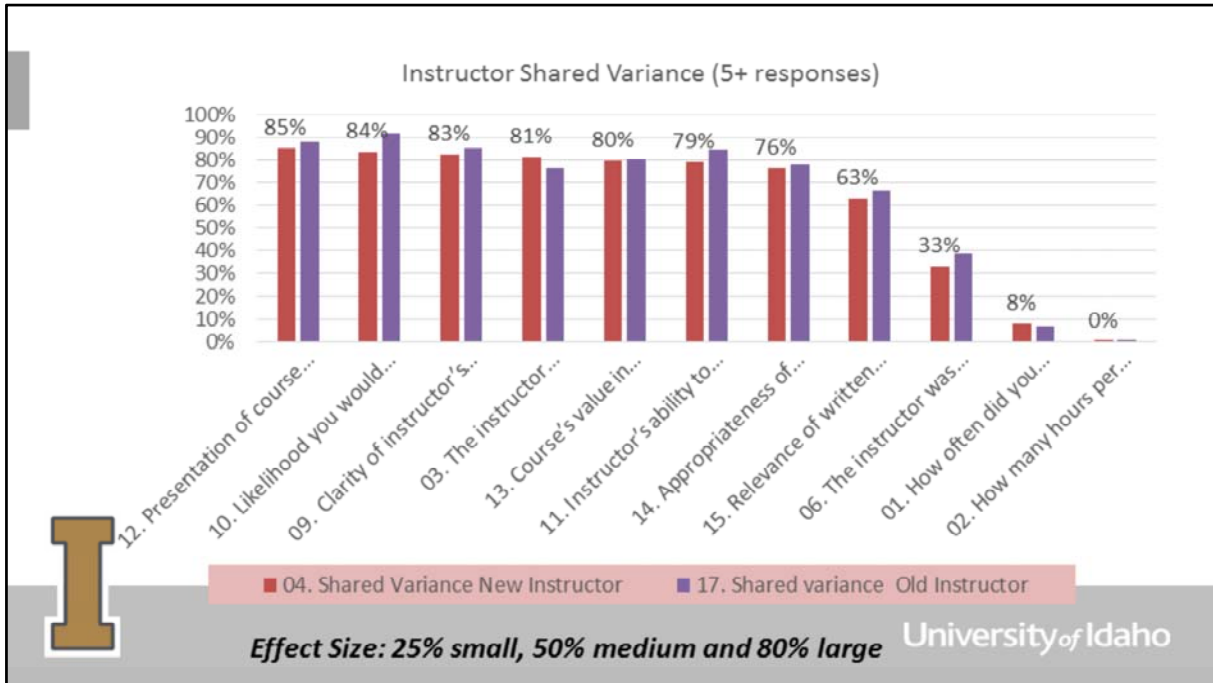


Effect Size  $r$ : 0.2 small, 0.5 medium and 0.8 large

University of Idaho

Discriminant Examination: The lowest three correlations from the administered items are shown here. There is strong consistency between the new and old instructor item.

Instructor behavior outside of class and student actions were the lowest. Student action items had very little influence and accounted for at most 8% of the shared variance (92% unaccounted for). The small positive relationship (8% shared variance) suggests that students that attended classes answered the instructor item slightly higher. The student rationale is unclear. It could be something like disliking the instructor's teaching style and therefore not going to class or because of greater exposure to the instructor. However, the strength of this relationship does not suggest it negates the utility of the item for its intended use.



This slide presents the shared variance of the items and the instructor scores (both old and new) . The general relationships are very similar and support the understanding of the instructor item presented earlier.

We see the most influential elements related to increased teacher scores were: 1) The presentation of the course material. 2) The instructor providing clarity and clear expectations for the course. 3) The course's value in gaining an understanding of the subject. 4) The likelihood that they would recommend the instructor to others (in the context of the elements described here); and 5) The instructor's ability to stimulate interest.



## INSTRUCTOR ITEM RELATIONSHIPS

<b>Correlations with Items to Old and New Instructor Scores: Top 5 (All Courses)</b>	<b>5. New Inst.</b>	<i>Rank New Inst.</i>	<b>17. Old Inst.</b>	<i>Rank Old Inst.</i>
03. The instructor expressed clear expectations for learning outcomes in this course.	0.87	01	0.81	07
12. Presentation of course material by the instructor.	0.86	02	0.90	02
10. Likelihood you would recommend this instructor to others.	0.85	03	0.93	01
09. Clarity of instructor's explanations.	0.85	04	0.89	03
11. Instructor's ability to stimulate interest in the course topics.	0.83	05	0.89	04
13. Course's value in gaining an understanding of the subject matter.	0.83	06	0.85	05



*Effect Size: 0.2 small, 0.5 medium and 0.8 large*

University of Idaho

Adding the low response classes to the data, as expected, lowered the relationships. However, it shows the same general pattern as the previous table. This supports reporting scores for courses with 5 or more responses.

## INSTRUCTOR ITEM RELATIONSHIPS

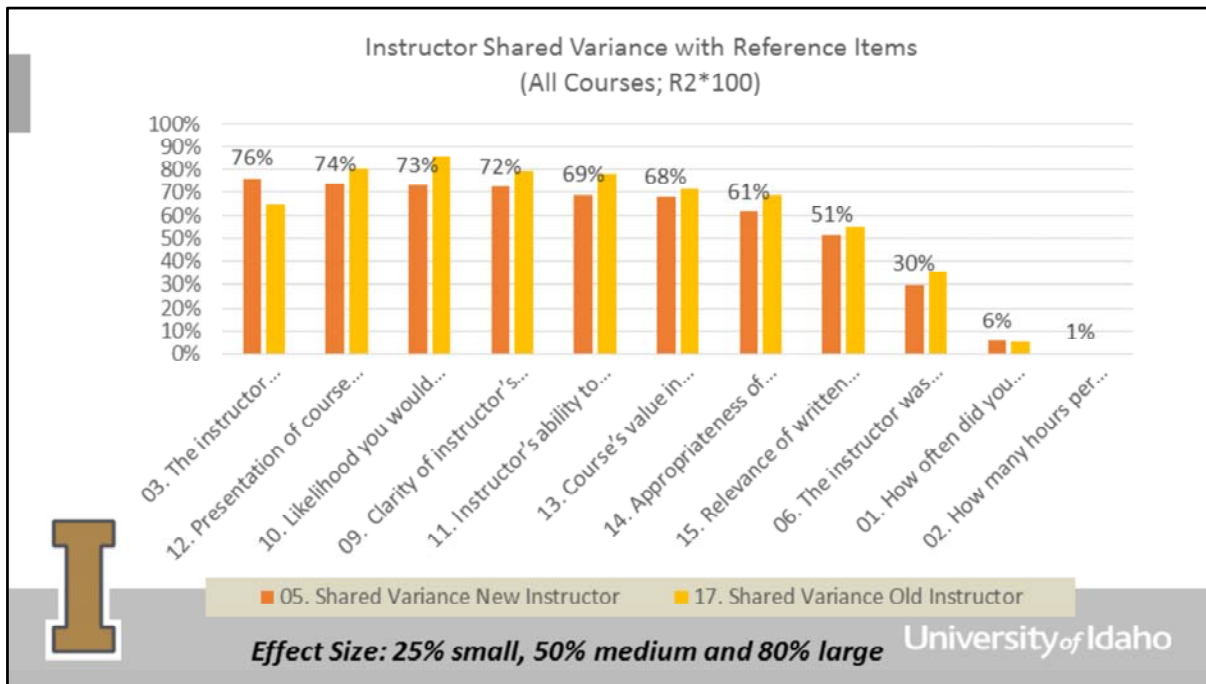
Correlations with Items to Old and New Instructor Scores: Lowest (All Courses)	5. New Inst.	Rank New Inst.	17. Old inst.	Rank Old Inst.
06. The instructor was helpful to me outside of class or in the online learning environment.	0.50	09	0.54	09
01. How often did you attend class or an online learning environment?	0.23	10	0.22	10
02. How many hours per week (outside of class) did you do work for this course?	0.08	11	0.06	11



Effect Size  $r$ : 0.2 small, 0.5 medium and 0.8 large

University of Idaho

The results, when including all courses, supports the reporting of courses with 5 or more responses and not small response scores.



The results, when including all courses, supports the reporting of courses with 5 or more responses and not small response scores.



## RELATIONSHIP WITH INSTRUCTOR DEMOGRAPHICS

Instructor and Course Items



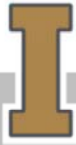
The operating characters of the grading structures at the graduate level typically functions differently from the undergraduate population. At the graduate level, the grading structure overall tends to be more of a pass-fail system.

In this analysis the undergraduate courses were merged with grades given in courses. Scores that are used with the DFWI reporting done at UI were used here. Those grades were used to create Course GPA (CGPA), DFWI percent and Percent A's scores. CGPA is calculated much as a student GPA by weighting grades given in a course by grade given. The total using only the graded students is averaged using a system where A=4, B=3, C=2, D=1 and F=0. This weighted average is the CGPA. The DFWI percent is a widely used metric where the percent of the total course enrollment who earned a grade of D, F W or I is calculated. The Percent A's score is the percent of graded students in a class that earned an A.

## INSTRUCTOR COURSE RESPONSES 5 OR MORE

Instructor Correlation Responses 5 or more				
	Q04 New Course	Q05 New Instructor	Q16 Old Course	Q17 Old Instructor
Responses	-0.06	-0.07	-0.06	-0.06
Enrollment	-0.06	-0.09	-0.08	-0.08
Response Rate	0.05	0.06	0.06	0.06

Instructor Shared Variance Responses 5 or more				
	Q04 New Course	Q05 New Instructor	Q16 Old Course	Q17 Old Instructor
Responses	0.3%	0.5%	0.4%	0.4%
Enrollment	0.4%	0.8%	0.7%	0.7%
Response Rate	0.2%	0.4%	0.4%	0.3%



*Correlation Effect Size r: 0.2 small, 0.5 medium and 0.8 large* University of Idaho

There were no meaningful relationship between responses, enrollments and response rates, and instructor or course scores. While there were small relationships between total responses, response rates and response rates, and the instructor and course scores, they did not have a practical effect. These relationships accounted for less than 1% of the variance (at most 0.8%). Here we see a slight lowering of the scores and enrollments as responses go up. There is also a slight increase in scores as the number of responses go up. These results support that for courses with 5 or more respondents, the scores are generally representative.

As total responses and enrollment are comparable in magnitude and in the negative direction while the response rate is positive, a similar magnitude suggests the decline would be due to class size. Bigger classes tend to have slightly lower scores. As the response rate increases, there is a slight increase in scores. Again, these data suggest these relationships have effect on the score. However, it is expected that there be more interaction with the instructor in smaller classes.

## INSTRUCTOR ALL COURSES

Instructor Correlation (All courses)				
	Q04 New Course	Q05 New Instructor	Q16 Old Course	Q17 Old Instructor
Responses	-0.09	-0.10	-0.10	-0.09
Class Size	-0.11	-0.13	-0.13	-0.12
Response Rate	0.10	0.10	0.11	0.10

Instructor Shared Variances (All courses)				
	Q04 New Course	Q05 New Instructor	Q16 Old Course	Q17 Old Instructor
Responses	0.9%	1.1%	0.9%	0.8%
Class Size	1.3%	1.7%	1.6%	1.3%
Response Rate	1.0%	1.1%	1.1%	1.0%



*Effect Size r: 0.2 small, 0.5 medium and 0.8 large*  
*Effect Size: 25% small, 50% medium and 80% large*

University of Idaho

## INSTRUCTOR GENDER AND ETHNICITY (5+ RESPONDENTS)

Instructor Correlation (All Courses)				
	Q04 New Course	Q05 New Instructor	Q16 Old Course	Q17 Old Instructor
Gender	-0.01	-0.01	-0.01	-0.02
Citizenship	-0.12	-0.12	-0.13	-0.13
Minority	-0.01	0.03	0.00	0.02

Instructor Shared Variance (All Courses)				
	Q04 New Course	Q05 New Instructor	Q16 Old Course	Q17 Old Instructor
Gender	0.0%	0.0%	0.0%	0.0%
Citizenship	1.4%	1.3%	1.8%	1.6%
Minority	0.0%	0.1%	0.0%	0.0%



*Effect Size r: 0.2 small, 0.5 medium and 0.8 large*  
*Effect Size: 25% small, 50% medium and 80% large*

University of Idaho

There was no meaningful relationship between citizenship, gender and ethnicity and the teaching and course scores. While the relationship of citizenship and course scores appears, it has little effect on scores by accounting for less than 1.5% of the variance of the new teaching or course score. Citizenship is coded as 0= US Citizen and 1= Not.

## INSTRUCTOR GENDER AND ETHNICITY (ALL COURSES)

Instructor Correlation (All Courses)				
	Q04 New Course	Q05 New Instructor	Q16 Old Course	Q17 Old Instructor
Gender	-0.02	-0.01	-0.02	-0.02
Citizenship	-0.07	-0.08	-0.09	-0.09
Minority	-0.01	0.01	0.02	0.01

Instructor Shared Variance (All Courses)				
	Q04 New Course	Q05 New Instructor	Q16 Old Course	Q17 Old Instructor
Gender	0.0%	0.0%	0.1%	0.0%
Citizenship	0.5%	0.6%	0.8%	0.9%
Minority	0.0%	0.0%	0.0%	0.0%



*Effect Size r: 0.2 small, 0.5 medium and 0.8 large*  
*Effect Size: 25% small, 50% medium and 80% large*

University of Idaho



## RELATIONSHIP WITH ACADEMIC INDICATORS

Instructor and Course Items



University of Idaho

The operating characters of the grading structures at the graduate level typically functions differently from the undergraduate population. At the graduate level, the grading structure overall tends to be more of a pass-fail system.

In this analysis the undergraduate courses were merged with grades given in courses. Scores that are used with the DFWI reporting done at UI were used here. Those grades were used to create Course GPA (CGPA), DFWI percent and Percent A's scores. CGPA is calculated much as a student GPA by weighting grades give in in a course by grade give. The total using only the graded students is average using a system where A=4, B=3, C=2, D=1 and F=0. This weighted average is the CGPA. The DFWI percent is a widely-used metric where the percent of the total course enrollment who earned a grade of D, F W or I is calculated. The Percent A's score is the percent of graded students in a class that earned an A.

## SHARED VARIANCE COURSE (5 + RESPONSES)

Academic Correlations N = 1827 (5 or more responses)			
	CRS GPA	% DFWI	% A's
Q04 New Course	0.23	-0.20	0.21
Q05 New Instructor	0.26	-0.23	0.24
Q16 Old Course	0.27	-0.26	0.24
Q17 Old Instructor	0.27	-0.24	0.26

Shared Variance (5 or more responses)			
	CRS GPA	% DFWI	% A's
Q04 New Course	5.1%	4.2%	4.3%
Q05 New Instructor	6.7%	5.1%	5.9%
Q16 Old Course	7.1%	6.6%	5.8%
Q17 Old Instructor	7.4%	5.7%	6.6%



*Effect Size r: 0.2 small, 0.5 medium and 0.8 large*  
*Effect Size: 25% small, 50% medium and 80% large*

University of Idaho

For courses with 5 or more respondents we find the outcome consistent with the literature in this area. As expected, we see the correlations in the .2 to .3 range. There is a small positive relationship in the expected direction between course and instructor scores. This is theoretically consistent as there is expected to be some relationship. Instructors who are experienced by students as assisting with their learning should be somewhat higher performing as a result. This theory assumes that instructors are value-added for the student. However, the literature in this area suggests that relationship should fall in the 0.20 to 0.30 range. This would mean that at most we would expect 9% or less of the variance to be accounted for in this case.

The correlations fall within the expected range and they account for at most 7.4% of the variance. The new questions consistently have a slightly smaller relationship accounting for less than 7% of the variance at most.

## SHARED VARIANCE COURSE (ALL COURSES)

Academic Correlation (All Courses)			
	CRS GPA	% DFWI	% A's
Q04 New Course	0.24	-0.20	0.23
Q05 New Instructor	0.27	-0.21	0.26
Q16 Old Course	0.29	-0.24	0.27
Q17 Old Instructor	0.28	-0.23	0.27

Shared Variance (All Courses)			
	CRS GPA	% DFWI	% A's
Q04 New Course	5.9%	4.1%	5.4%
Q05 New Instructor	7.2%	4.6%	6.8%
Q16 Old Course	8.3%	5.8%	7.6%
Q17 Old Instructor	7.7%	5.4%	7.2%



*Effect Size r: 0.2 small, 0.5 medium and 0.8 large*  
*Effect Size: 25% small, 50% medium and 80% large*

University of Idaho

For courses with 5 or more respondents we find the outcome consistent with the literature in this area. As expected, we see the correlations in the .2 to .3 range. There is a small positive relationship in the expected direction between course and instructor scores. This is theoretically consistent as there is expected to be some relationship. Instructors who are experienced by students as assisting with their learning should be somewhat higher performing as a result. This theory assumes that instructors are value-added for the student. However, the literature in this area suggests that relationship should fall in the 0.20 to 0.30 range. This would mean that at most we would expect 9% or less of the variance to be accounted for in this case.

The correlations fall within the expected range and they account for at most 7.4% of the variance. The new questions consistently have a slightly smaller relationship accounting for less than 7% of the variance at most.

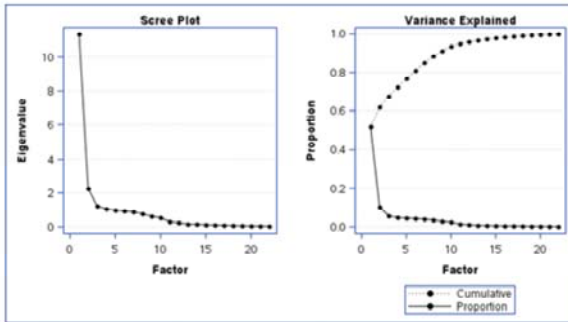
EXPLORATORY FACTOR STRUCTURE OF  
INSTRUCTOR AND COURSE ITEMS, ADDITIONAL  
ITEMS, AND INSTRUCTOR DEMOGRAPHICS  
ACADEMIC INDICATORS

Instructor and Course Items



University of Idaho

# INFORMATION ON FACTOR STRUCTURE



Inter-Factor Correlations				
	Factor1	Factor2	Factor3	Factor4
Factor1	1.00			
Factor2	0.21	1.00		
Factor3	0.08	0.04	1.00	
Factor4	-0.06	0.01	0.05	1.00

The 4 factor model was theoretically consistent and accounted for 72% of the variance. This model clearly differentiates instructor items from academic and the other elements, and has a correlation of 0.21 between those factor (F1 & F2).



University of Idaho

Using the same sample of courses used to examine the relationship to academic measures so as to include the academic indicators along with all items while including the faculty demographics a PCA analysis with an oblique rotation was performed. The oblique rotation allows the various underlying factors to be correlated with one another.

Various factor structures were examined ranging from 3 to 10 factors using the scree and eigenvalue rule of one as a general guide. The 10 factor solution was examined as it was the point where the proportion of variance accounted for began to plateau. The 4 factors solution fitted with the eigenvalue rule of 1 in place was most parsimonious. The increase in factors does not blur the 1st factor at all, and results in the other factors become distinct factors. The only change in factor 1 across the various increased numbers of factors is that the item addressing the helpfulness of the instructor outside of class breaks out into a separate factor by the time we get to the 10 factor solution.

The 4 factor solution accounts for about 72% of the variance. The items per factor appears next.

\*\*DALE: Sara's edit (first line) is in red, and I changed the font so it's easier for you to find.

\*\*\*FYI: I don't have a clue what this note says...no idea.

## OBLIQUE FACTORING (PCA) 5 OR MORE RESPONSES

Factor 1: Instructor/Course Elements (Educational Atmosphere)	Loading
17. (Old) Overall, how would you rate the instructor's performance in teaching this course?	0.97
16. (Old) Overall, how would you rate the quality of this course?	0.96
12. Presentation of course material by the instructor.	0.96
05. (New) Overall, the instructor's delivery and efforts contributed to your understanding of the course material.	0.96
09. Clarity of instructor's explanations.	0.95
13. Course's value in gaining an understanding of the subject matter.	0.95
04. (New) Overall, the content and organization of this course contributed to your understanding of this subject	0.95
10. Likelihood you would recommend this instructor to others.	0.94
14. Appropriateness of level at which course material is covered.	0.94
03. The instructor expressed clear expectations for learning outcomes in this course	0.93
11. Instructor's ability to stimulate interest in the course topics	0.93
15. Relevance of written assignments to course materials.	0.87
06. The instructor was helpful to me outside of class or online learning environment.	0.63

Factor 1 loadings are presented here in decreasing magnitude. The majority of the items load essentially equivalently. A course item and an items-related student-instructor interaction outside the classroom were somewhat lower loading. The general theme of these items appears to be instructor-controlled aspects representing the of educational atmosphere of the course.

## OBLIQUE FACTORING (PCA) 5 OR MORE RESPONSES

Factor 2: Academics and Focus (Student)	Loading
22. CRS GPA	0.91
24. PCT A's	0.90
23. PCT DFWI	-0.75
02. How many hours per week (outside of class) did you do work for this course?	-0.34



University of Idaho

Factor 2 loading are presented here in decreasing magnitude. The academic measures group together in the expected direction with a loading of student effort for the course, which is expected. The student loading suggests that if students do not put time in outside the course and only do work in the course, they are more likely to withdraw. As student effort interacts with skill and ability, it is not unreasonable that this relationship would be lower.

## OBLIQUE FACTORING (PCA) 5 OR MORE RESPONSES

Factor 3: Student Effort and Attendance (Interest or Focus)	Loading
02. How many hours per week (outside of class) did you do work for this course?	0.65
18. Number of student reponses	-0.60
01. How often did you attend class or online learning environment?	0.58



University of Idaho

Factor 3 loading are presented here in decreasing magnitude. The items in this factor essentially group around the theme of student attendance and effort or focus. If they do not spend much time studying they are more likely to not be in class. In that case, they are also less likely to complete the inventory. The study time item here loads as little over twice as high here as it did on the previous factor. It is the only item that loads cross-across 2 or more factors.



## OBLIQUE FACTORING (PCA) 5 OR MORE RESPONSES

Factor 4: Instructor Demographics	Loading
21. Minority (0 Minority 1 White or UNK)	0.66
19. Gender (0 Male 1 Female)	0.63
20. Citizenship (0 Citizen 1 Not)	0.51



University of Idaho

Factor 4 loadings are presented here in decreasing magnitude. These are all Instructor Demographics. These are point bi-serial relationships in the model.

## COURSE ITEMS: RELATIONSHIPS TO OTHER ITEMS AND DEMOGRAPHICS

- 4. (New) Overall, the content and organization of this course contributed to your understanding of this subject
- 16. (Old) Overall, how would you rate the quality of this course?



University of Idaho

Here we can examine the wording of the new item and older course item in more detail. Here the new course item incorporates some instructor-related elements such as organization. As previously noted, the course and instructor items do not differentiate from one another well.

As course relationship to academics, instructor demographics, and the factor structure were examined previously, this section will focus on the course items relationship to the additional items.

## ADDITIONAL ITEMS

01. How often did you attend class or online learning environment?
02. How many hours per week (outside of class) did you do work for this course?
03. The instructor expressed clear expectations for learning outcomes in this course.
06. The instructor was helpful to me outside of class or online learning environment.
09. Clarity of instructor's explanations
10. Likelihood you would recommend this instructor to others
11. Instructor's ability to stimulate interest in the course topics
12. Presentation of course material by the instructor
13. Course's value in gaining an understanding of the subject matter
14. Appropriateness of level at which course material is covered
15. Relevance of written assignments to course materials



University of Idaho

Here we can see the list of additional items that are examined in relationship to the course questions. The relationships reported here are Pearson correlations, except for item 6 which is a point-biserial correlation as it is binary (Yes/No with NA's excluded). Spearman correlations were examined in most cases as well, with the results being similar. In some cases they were slightly lower and others slightly higher. There was little difference and as the professional literature in this area typically reports Pearson and point-biserial, that is what is being reported here.

## COURSE ITEM RELATIONSHIPS

Correlations with Items Selected to Compare Old and New Course Scores Top 5 (Responses 5 or more)	4. New Course	Rank New Course	16. Old course	Rank Old Course
13. Course's value in gaining an understanding of the subject matter.	0.91	1.00	0.93	1.00
03. The instructor expressed clear expectations for learning outcomes in this course	0.91	2.00	0.86	7.00
12. Presentation of course material by the instructor.	0.88	3.00	0.90	4.00
14. Appropriateness of level at which course material is covered.	0.88	4.00	0.92	3.00
09. Clarity of instructor's explanations.	0.88	5.00	0.89	6.00
10. Likelihood you would recommend this instructor to others.	0.87	6.00	0.92	2.00
11. Instructor's ability to stimulate interest in the course topics	0.85	7.00	0.90	5.00

Effect Size: 0.2 small, 0.5 medium and 0.8 large

University of Idaho

Convergent Examination: The relationships ranged from 0.91 to -0.01. The top five relationships for the new and old course items with the additional items are shown. These represent the upper limit of the relationship as the overlap of the variance with other items was not removed. This range suggests that students responded differently to the items (not just marking all the same). However, due to the limited number of items used here, more investigation on this would be needed for the new item. Given the correspondence of the new item with the old item, it seems reasonable these results suggest that there will be similar patterns as noted for the old item. Previous research with the old item data provided further support for the differentiated item responses by students.

Here we see the more influential elements related to increased course scores were: 1) The course value in understanding the topic and clear expectations of the instructor. 2) The presentation style of the instructor. 3) The appropriateness of the level of course material (not teaching over their heads); and, 4) The clarity of the instructor's explanations. These data suggest that while course elements are present, it is heavily loaded with instructor elements, perhaps more so than the old item. However, as the old Course and Instructor scores show little discrimination, there is little suggested change when using the new items.

Several multiple regression models were employed to look at the variance accounted for with the New Course item. The model with the first 5 items accounted for 89% of the variance. Various combinations created by adding any other variables from the validity items and demographic variables did not increase the variance accounted for over this model. The full model with all of these variables stayed at 89%.

## INSTRUCTOR ITEM RELATIONSHIPS

Correlations with Items to Old and New Course Scores: Lowest (5+ responses)	4. New Course	Rank New Course	16. Old course	Rank Old Course
06. The instructor was helpful to me outside of class or in the online learning environment.	0.55	9	0.58	9
01. How often did you attend class or online learning environment?	0.28	10	0.26	10
02. How many hours per week (outside of class) did you do work for this course?	-0.01	11	-0.03	11

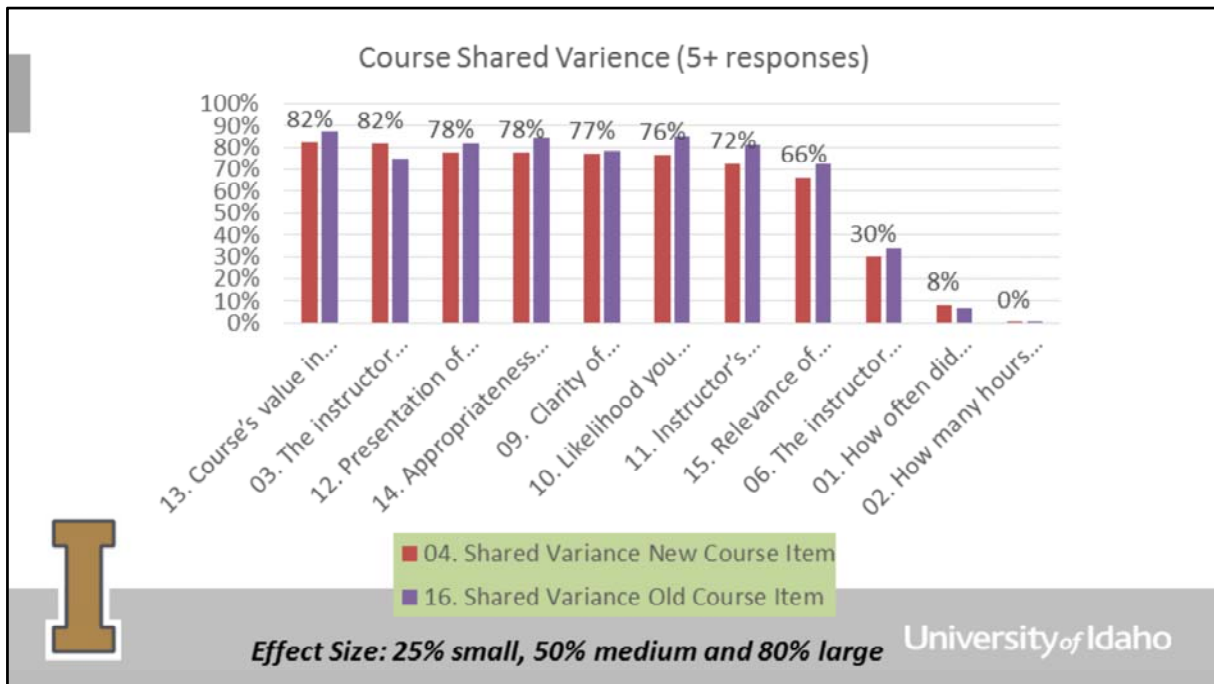


*Effect Size r: 0.2 small, 0.5 medium and 0.8 large*

University of Idaho

Discriminant Examination: The lowest three correlations from the administered items are shown here. There is strong consistency with the new and old course item.

Instructor behavior outside of class and student actions were the lowest (30% shared variance). Student action items had very little influence and accounted at most for 8% of the shared variance. The small positive relationship (8% shared variance) suggests that students who attended classes more frequently answered the instructor item slightly higher. What the student motivation was is unclear. It could be things like disliking the instructor's teaching style and therefore not going to class, or working more hours, etc. However, the strength of this relationship does not suggest it negates the utility of the item for its intended use.



This slide presents the shared variance of the items and the course scores (both old and new) . The general relationships are very similar and support the understanding of the course item presented earlier.

We see the most influential elements related to increased course scores were: 1) The course's value in gaining an understanding of the subject. 2) The instructor providing clarity and clear expectations for the course. 3) The presentation of the course material; and, 4) the appropriateness of the level of instruction (not talking over their heads, etc.).

## SUMMARY

- The New Course and Instructor items perform equivalently to the older items.
- The items appear to be largely unaffected by the demographic measures, enrollment and response rate.
  - Slight decrease of scores for larger classes and slight increase for more responses and increased response rate.
- There is little discrimination between course and instructor score with both the old and new course and items scores.
- These data suggest there would be reasonable continuity if the transition is made relative to promotion and tenure.





University of Idaho