

TO: CIOS Task Force Members

FROM: Dr. Tris Utschig
Assistant Director for the Scholarship and Assessment of Teaching and Learning
Center for the Enhancement of Teaching and Learning
Georgia Institute of Technology

Sunni Newton
Graduate Research Assistant
Center for the Enhancement of Teaching and Learning
Georgia Institute of Technology

DATE: September 27, 2010

RE: **SU10 Digital Measures PILOT Survey Summary**

Dear Members of the CIOS Task Force,

In accordance with the recommendation of the Task Force, the Executive Board, and the Associate Dean's Council, CETL conducted a PILOT survey in the Summer of 2010 to support ongoing investigations into means for improving the CIOS instrument. This follow up PILOT focused on the use of an external vendor as the means to deliver the CIOS survey. The content of the survey is the same as that of the Fall 2008 PILOT. Attached is a summary of the results gathered from the PILOT survey using the external vendor Digital Measures, along with summary data from the combined results of the SU08, FA08, and SU10 PILOT efforts. These results will support discussion about how to move forward with the CIOS instrument.

Respectfully submitted,
Dr. Tris Utschig & Sunni Newton

Summer 2010 Digital Measures (DM) PILOT Summary, Analysis, & Comparison with CIOS

Executive Summary

1,426 total responses (615 from CIOS, 811 from DM PILOT) representing 25 different courses with 40 individual course sections, 10 departments, and 17 instructors are included in the analysis.

The manner in which items functioned on the two instruments was assessed by comparing item responses for 6 sets of similar items and 1 identical item. The identical item (overall effectiveness) functioned similarly in the two instruments. Three of the six sets of similar items functioned similarly (clarity, affect, and encouragement), indicating that a switch to the DM PILOT would seem to provide similar data to the CIOS in these areas. Three sets functioned somewhat dissimilarly (workload, content level, and relevance). In general, data from the three PILOT vs. CIOS comparisons (summer 2008, fall 2008, and summer 2010) reveals that, for the most part, students provide roughly comparable responses on similar items from CIOS and PILOT. Few courses showed large discrepancies between similar items, and students typically did not display strong tendencies to provide higher scores on one item or the other within a given set of similar items. Exceptions to the general findings will be discussed below.

Implementation

1,426 total responses (615 from CIOS, 811 from DM PILOT) responses representing 25 different courses with 40 individual course sections, 10 departments, and 17 instructors are included in the analysis. Both the CIOS and DM PILOT were anonymous and thus only an extremely limited comparative analysis was possible. The items from each survey are presented in Appendix A.

Response Rate

Summer response rates are generally relatively low compared to typical response rates for fall and spring, averaging roughly 25% for the summer of 2010. Response rates for courses utilizing the DM PILOT ranged from 23 to 100%, with an average of 64%; for this same set of courses, response rates for CIOS ranged from 4 to 100% with an average of 49%. So response rates were slightly higher for the DM PILOT as compared to the regular CIOS in this set of courses, and response rates on both the DM PILOT and CIOS for this set of courses were higher than the overall CIOS response rate for all courses. Response rates on each survey and enrollment for each course are provided in Table 1.

Table 1. Response rates and course enrollments.

Course	Course Enrollment (per CIOS)	CIOS: # responses	DM PILOT: # responses	CIOS: response rate	DM PILOT: response rate
1	26	1	6	4%	23%
2a	16	7	16	44%	100%
2b	16	6	16	38%	100%
3a	2	1	2	50%	100%
3b	2	1	2	50%	100%
4	28	22	26	79%	93%
5	16	16	16	100%	100%
6	16	16	16	100%	100%
7	85	71	49	84%	58%
8	29	3	13	10%	45%
9	26	4	26	15%	100%
10	12	9	12	75%	100%
11	36	3	14	8%	39%
12	25	6	11	24%	44%
13	2	1	1	50%	50%
14	48	38	41	79%	85%
15	35	32	31	91%	89%
16	13	10	9	77%	69%
17	13	3	9	23%	69%
18	44	6	20	14%	45%
19	2	1	2	50%	100%
20	26	15	17	58%	65%
21	26	21	20	81%	77%
22	24	16	21	67%	88%
23	23	15	16	65%	70%
24	25	21	18	84%	72%
25	25	17	19	68%	76%
26	18	7	10	39%	56%
27	25	18	18	72%	72%
28	59	21	42	36%	71%
29	35	12	14	34%	40%
30	10	1	8	10%	80%
31	30	9	13	30%	43%
32	23	12	13	52%	57%

33	11	4	4	36%	36%
34	74	28	47	38%	64%
35	56	30	29	54%	52%
36	30	11	26	37%	87%
37	99	39	46	39%	46%
38	122	51	73	42%	60%
39	24	7	13	29%	54%
40	4	3	4	75%	100%

Comparison of findings from three CIOS vs. PILOT studies (summer 2008, fall 2008, and summer 2010).

One small (summer 2008) and two larger (fall 2008, summer 2010) studies were run in which students from a sample of courses were asked to fill out both the CIOS and the PILOT for these courses. Collecting data with both instruments from students in the same course allowed for a review of how similarly, or dissimilarly, the two instruments functioned. For the most part, comparable items on the two surveys functioned similarly.

Data from these three comparison studies are presented in Table 2. Interpolated median scores for each item within a set of similar items were compared for a given course, and the courses were sorted based on the size of this difference for each set of similar items. Courses that showed a difference in interpolated median scores on a given set of items are also broken down, by percentage, based on whether they displayed higher interpolated median scores on the CIOS or on the PILOT.

Table 2. Data from CIOS vs. PILOT comparisons for summer 2008, fall 2008, and summer 2010

	Clarity Items			Affect Items			Encouragement Items		
	Summer 2008 (N = 16)	Fall 2008 (N = 50)	Summer 2010 (N = 42)	Summer 2008 (N = 16)	Fall 2008 (N = 50)	Summer 2010 (N = 42)	Summer 2008 (N = 16)	Fall 2008 (N = 50)	Summer 2010 (N = 42)
No difference:	13%	18%	21%	13%	32%	26%	n/a	22%	26%
Difference = 0.1 – 0.2:	25%	44%	43%	31%	40%	40%	n/a	46%	38%
Difference = 0.3 – 0.5:	38%	36%	21%	19%	22%	24%	n/a	22%	24%
Difference = 0.6 – 1.0:	25%	2%	10%	31%	6%	10%	n/a	10%	12%
Difference > 1.0:	0%	0%	5%	6%	0%	0%	n/a	0%	0%
Difference of .5 or less:	75%	98%	85%	63%	94%	90%	n/a	90%	88%
CIOS higher	64%	32%	39%	29%	38%	58%	n/a	46%	48%
PILOT higher	36%	68%	61%	71%	62%	42%	n/a	54%	52%

	Relevance Items			Content Level Items			Overall Effectiveness Item		
	Summer 2008 (N = 16)	Fall 2008 (N = 50)	Summer 2010 (N = 42)	Summer 2008 (N = 16)	Fall 2008 (N = 50)	Summer 2010 (N = 42)	Summer 2008 (N = 16)	Fall 2008 (N = 50)	Summer 2010 (N = 42)
No difference:	47%	6%	23%	n/a	16%	18%	50%	30%	33%
Difference = 0.1 – 0.2:	20%	26%	33%	n/a	35%	33%	13%	40%	40%
Difference = 0.3 – 0.5:	13%	42%	28%	n/a	35%	38%	13%	24%	14%
Difference = 0.6 – 1.0:	20%	26%	15%	n/a	12%	8%	25%	6%	10%
Difference > 1.0:	0%	0%	0%	n/a	2%	3%	0%	0%	2%
Difference of .5 or less:	80%	74%	84%	n/a	86%	89%	75%	94%	87%
CIOS higher	88%	85%	66%	n/a	46%	53%	25%	46%	43%
PILOT higher	13%	15%	33%	n/a	54%	47%	75%	54%	57%

In all cases the majority, and in most cases at least 75%, of the course sections showed a ½ point or smaller difference in interpolated median scores on a set of comparable items. In terms of whether students provided higher ratings on the CIOS or the PILOT item within a set of comparable items, the breakdown was close to 50/50 for the content level items and for the encouragement items, suggesting that students are roughly equally likely to provide a higher score on the CIOS as on the PILOT item for these two sets of items. This is also the case for the overall effectiveness item except in the case of the summer 2008 data; we believe that the large discrepancy shown in the summer 2008 data, in which 75% of courses with a difference provided a higher rating on the PILOT, may be less conclusive given the very low sample size on which this percentage breakdown is based (i.e., only 8 course sections).

For the clarity and affect items the trend of providing higher scores on the CIOS or on the PILOT fluctuated, such that the data, when viewed over the three semesters, do not provide evidence for a clear trend of higher scores on one instrument or the other. The relevance items do show a clear trend for students providing a higher score on the CIOS than on the PILOT item. This difference is likely due to changes in the nature of the items. One of these items pertains to exams and quizzes (CIOS) while the other deals with assignments and activities (PILOT). It seems reasonable that some students may feel quite differently about the assignments and the evaluations they were given in a course. Furthermore, one item asks about covering course

objectives and content (CIOS) while the other asks about facilitating learning (PILOT). Although a strong relationship between these two ideas should be expected, it is possible that students conceptualized these two evaluative components differently.

Lastly, a comparison was made between two items related to workload which is different than the comparisons for the other sets of items. The PILOT workload item asks for a report of the number of hours worked per week for the course while the CIOS item asks about the appropriateness of the number of course assignments. A correlation was conducted between these two items to assess their relationship. For the fall 2008 and summer 2010 data, negative relationships were found between these two items ($r = -.29$ and $r = -.50$, respectively), such that reports of more hours worked per week were related to lower ratings of the appropriateness of the number of assignments. Conversely, a positive relationship was found between these two items for the summer 2008 data ($r = .26$). This may be due to something specific about the set of courses included in the summer 2008 study (i.e., students in these courses found it beneficial to work more hours per week, or had expectations of working more hours per week, etc.). The summer 2008 study also had a much smaller set of courses, and as such is likely somewhat less conclusive than the data from the fall 2008 and summer 2010 studies.

Survey Data from Instructors Participating in Summer, 2010 DM PILOT vs. CIOS Study

Invitations to participate in a survey on perceptions of the DM PILOT vs. the CIOS were sent to the 21 instructors who participate in this study. Nine instructors responded to the survey for a response rate of 43%. These respondents answered several questions asking about whether they preferred the DM PILOT, preferred the CIOS, or thought the two surveys were equal with respect to several aspects of using the instruments. These data are presented in Table 3.

These data suggest that in all cases, a larger number of respondents reported preferring the DM PILOT rather than the CIOS. For each aspect of survey use that was evaluated, no more than one respondent reported a preference for the CIOS. The majority of instructors reported that the two instruments were equal in terms of ease of use in monitoring student participation during the survey period, ease of use in accessing reports, and potential to impact their teaching. The majority of instructors reported preferring the DM PILOT for usefulness in understanding student perceptions of their teaching and for contents of reports. In conclusion, instructors perceive the two instruments as being equivalent for several aspects of survey use, but for aspects of survey use on which a substantial number of respondents do report a preference, they tend to prefer the DM PILOT.

Table 2. Instructor Survey Data.

	DM is preferable	CIOS is preferable	both are equal	N/A
Which survey is more useful for understanding student perceptions of your teaching?	5	1	3	0
Which survey do you prefer for ease of use in choosing additional items to be included in the survey?	1	1	2	5
Which survey do you prefer for ease of use in monitoring student participation during the survey period?	3	0	5	1
Which survey do you prefer for ease of use in accessing reports?	2	0	7	0
Which survey do you prefer for contents of reports?	5	1	3	0
Which survey do you prefer for formatting of reports?	4	1	4	0
Which survey data has more potential to impact your teaching?	2	1	6	0

These instructors also provided some comments regarding their perceptions of the DM PILOT and CIOS and how these two instruments compare. In general the comments seem to suggest that most instructors preferred the questions included on the DM PILOT but had some difficulties in working with the reports generated from the DM PILOT. It also seems that instructors did not see a strong preference among their students for one survey or the other.

- Several instructors commented on their preference for the questions on the DM PILOT:
 - “[DM PILOT] Seems to solicit more in-depth verbal comments.”
 - “Some of the questions were better worded on the Digital Measures instrument.”
 - “The extra questions about teacher enthusiasm, etc. really provoked some interesting responses from the students. Because I taught two courses during the summer, I had excellent material for comparison.”
 - “The feedback to the Digital Measures survey is more direct and practical, probably as a result of the adapted formulation of the questions.”
 - “I do like the first set of questions [on DM PILOT] about students’ perception of their workload and participation.”
- Some instructors expressed a preference for the comments generated by CIOS as compared to Digital Measures:
 - “On CIOS my students wrote comments in addition to the numerical scoring. The comments are most helpful to me.”
 - “Written comments [on CIOS] more useful than numbers.”
- Instructors provided comments about the formatting of the DM PILOT and the reports generated from the DM PILOT data, some positive and others negative:
 - “I like the excel format [of the DM PILOT]”
 - “DM was much easier for generating various versions of my reports.”
 - “I don't like having to work with the formatting of the open-ended responses with the Digital Measures output. I would prefer it be formatted in an easy-to-read manner (not in a spreadsheet).”
 - “It took me time to reformat the spreadsheet data [from the DM PILOT] in order to view it easily. How about providing both a .csv file and a formatted text file with 2 or 3 responses summarized on each page.”
 - “I couldn't find the way to add additional questions or provide space for personal comments. Maybe this is possible, but I didn't see it.”
- Instructors indicated that their students did not express a strong preference for one survey or the other, but one instructor indicated that some students had technical difficulties accessing the DM PILOT:
 - “They said it was no trouble so I think they were indifferent between the two.”
 - “The students' response to my question about their experiences: "they are just the same". I was actually surprised to see the more detailed feedback on Digital Measures. Apparently, they did not perceive that as an extra, unreasonable burden.”
 - “Some students said it froze their machine.”

- **Comparison of Numerical Results for Similar Questions between CIOS/DM PILOT**

Summary of Findings

One identical item and six sets of similar items were compared to investigate differences between these items on the two survey instruments. For the identical item (overall effectiveness) and three sets of the similar items (clarity, affect, and encouragement), the scores were reasonably comparable for most courses, indicating that these items seem to be functioning similarly on the PILOT and CIOS. Three sets of items (workload, content level, and relevance), reflected slightly discrepant, although still roughly comparable, responses across the two surveys, indicating that they may be tapping into different student opinions. Item content on all items sets for both surveys has been analyzed previously; please see the January, 2009 Expanded PILOT report for a discussion of item content.

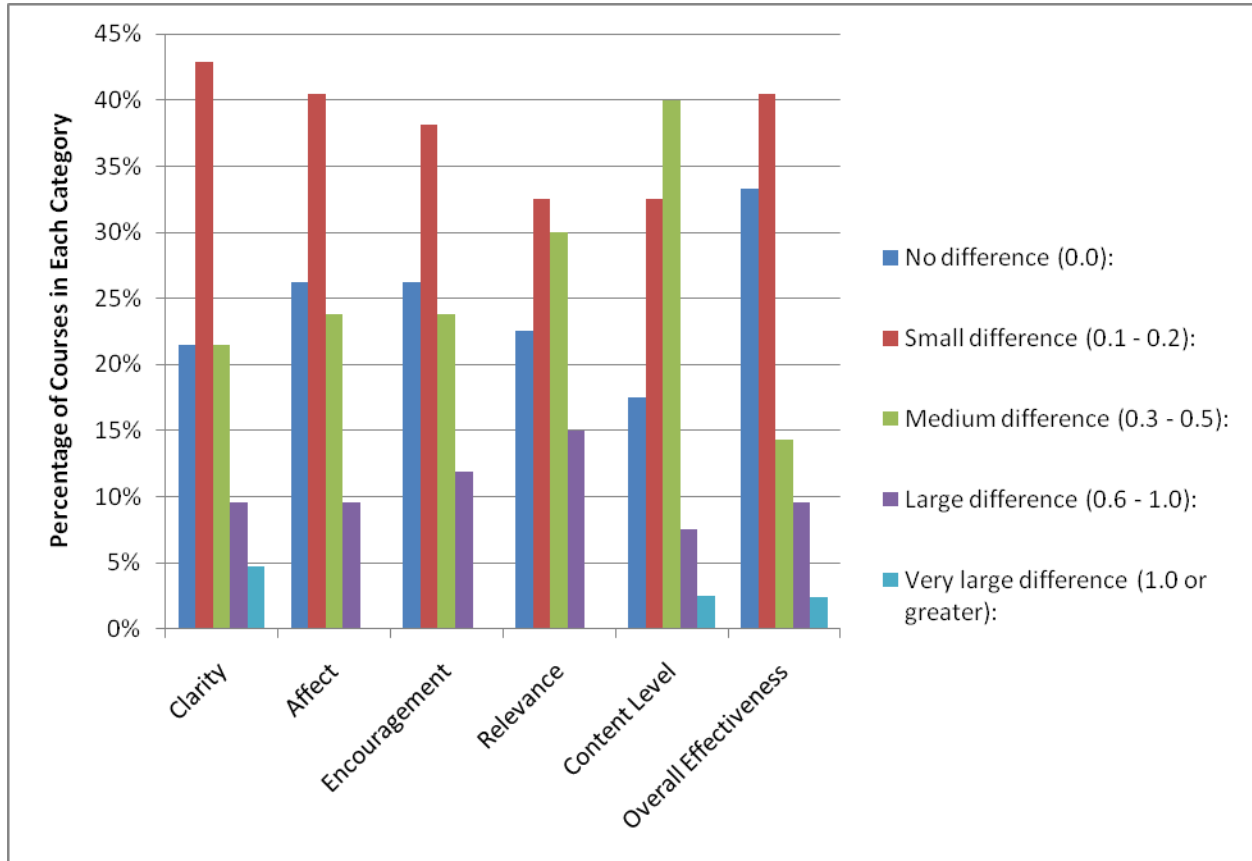
Detailed Analyses

One identical item (CIOS Item 10) and six sets of comparable items were included on the CIOS and DM PILOT surveys. A comparison of the numerical values for the interpolated medians on these items is displayed in Figure 1. For each item, course sections are categorized based on the magnitude of the difference between interpolated median ratings on the DM PILOT and CIOS (Difference = $|IM_{DM\ PILOT} - IM_{CIOS}|$). The percentage of courses in each of the following categories are reported.

- No difference
- Difference of 0.1 to 0.2 points
- Difference of 0.3 to 0.5 points
- Difference of 0.6 to 1.0 points
- Difference of greater than 1.0 point

Note: All comparisons for sets of items include 40 - 42 course sections.

Figure 1. DM PILOT and CIOS Similar Item Comparison.



Identical Item Comparison: Overall Effectiveness (CIOS item 10)

“Considering everything the instructor was an effective teacher.”

This question, commonly referred to as “Item 10” on the CIOS instrument, was duplicated verbatim on the DM PILOT (Quality of Teaching, Item 8).

Data Analysis

- Ratings were observed to be largely consistent between the two surveys, though not equivalent, with 31 out of the 42 courses (74%) for which a comparison was possible having a difference of 0.2 or smaller between the two surveys.
- There was no clear trend toward higher responses on this item in one survey or the other; for the 28 classes in which responses differed between the two surveys, 12 (43%) favored higher responses on the CIOS and 16 (57%) favored higher responses on the DM PILOT.
- The largest difference between interpolated medians on this item for any section was 0.9.
- Of the 42 classes for which a direct comparison was possible, the categorization of course sections based on size differences in the interpolated medians for the two items is as follows:

Item Comparison For Overall Effectiveness		
Difference Size Category:	Count:	Percentage:
No Difference:	14	33%
Difference = 0.1 - 0.2:	17	40%
Difference = 0.3 - 0.5:	6	14%
Difference = 0.5 - 1.0:	4	10%
Difference > 1.0:	1	2%

- Possible explanations for these differences include different students from a course completing the different surveys, the use of forced choice answers rather than continuous scales, and the influence of the content of the questions asked on the survey prior the Item 10 question.

Similar Item Comparison: Clarity

CIOS item 3:	“The instructor explained complex material clearly.”
DM PILOT item Quality of Teaching 1:	“Instructor’s clarity in discussing or presenting course material”

Data Analysis:

- In general, scores on these items were somewhat comparable, although not equivalent. While the majority of courses for which a comparison was possible had a difference equal to 0.2 or smaller (27 out of 42 courses or 64%), sizeable differences of 1.1 and 1.5 were seen in two of the courses. So it appears that in some cases, students are interpreting these items differently. Perhaps this may be driven by the difference in wording of the items, in that the DM PILOT item is more general.
- Of the 42 classes for which a direct comparison was possible, the categorization of course sections based on size differences in the interpolated medians for the two items is as follows:

Item Comparison For Clarity		
Difference Size Category:	Count:	Percentage:
No Difference:	9	21%
Difference = 0.1 - 0.2:	18	43%
Difference = 0.3 - 0.5:	9	21%
Difference = 0.5 - 1.0:	4	10%
Difference > 1.0:	2	5%

- The largest difference between interpolated medians on these items for any section was 1.5.

- There was a moderately strong trend towards higher scores on the DM PILOT item as compared to the CIOS item; for the 33 classes in which responses differed between the two surveys, 20 (61%) showed a higher interpolated median response on the DM PILOT item and 13 (39%) showed a higher interpolated median response on the CIOS item. Again, this difference may be driven by the differential wording of the DM PILOT item, which may prompt the student to consider a wider range of teacher behaviors and thus possibly form a more positive view of the instructor’s clarity.

Similar Item Comparison: Affect

CIOS item 4:	“The instructor was approachable and willing to assist individual students.”
DM PILOT item Quality of Teaching 3:	“Instructor’s respect and concern for students”

Data Analysis:

- In general, scores on these items were fairly comparable, although not equivalent, with the majority of courses having a difference of 0.2 or smaller between the two survey items (28 out of 42 courses, or 66%), and no courses having a difference of greater than 1.0.
- Of the 42 classes for which a direct comparison was possible, the categorization of course sections based on size differences in the interpolated medians for the two items is as follows:

Item Comparison For Affect		
Difference Size Category:	Count:	Percentage:
No Difference:	11	26%
Difference = 0.1 - 0.2:	17	40%
Difference = 0.3 - 0.5:	10	24%
Difference = 0.5 - 1.0:	4	10%
Difference > 1.0:	0	0%

- The largest difference between interpolated medians on these items for any section was 1.0.
- There was no clear trend towards higher scores on one survey item as compared to the other; for the 33 classes in which responses differed between the two surveys, 13 (42%) showed a higher interpolated median response on the DM PILOT item and 18 (58%) showed a higher interpolated median response on the CIOS item.

Similar Item Comparison: Encouragement

CIOS item 5:	“The instructor encouraged students to consult with him or her.”
DM PILOT item Quality of Teaching 6:	“The instructor was readily available for consultation.”

Data Analysis:

- In general, scores on these items were fairly comparable, although not equivalent, with the majority of courses having a difference between the two survey items of 0.2 or less (27 out of 42 courses, or 64%), and no courses having a difference greater than 1.0.
- Of the 42 classes for which a direct comparison was possible, the categorization of course sections based on size differences in the interpolated medians for the two items is as follows:

Item Comparison For Encouragement		
Difference Size Category:	Count:	Percentage:
No Difference:	11	26%
Difference = 0.1 - 0.2:	16	38%
Difference = 0.3 - 0.5:	10	24%
Difference = 0.5 - 1.0:	5	12%
Difference > 1.0:	0	0%

- The largest difference between interpolated medians on these items for any section was 0.7.
- There was no clear trend toward higher responses on this item in one survey or the other; for the 31 classes in which responses differed between the two surveys, 16 (52%) showed a higher interpolated median response on the DM PILOT item and 15 (48%) showed a higher interpolated median response on the CIOS item.

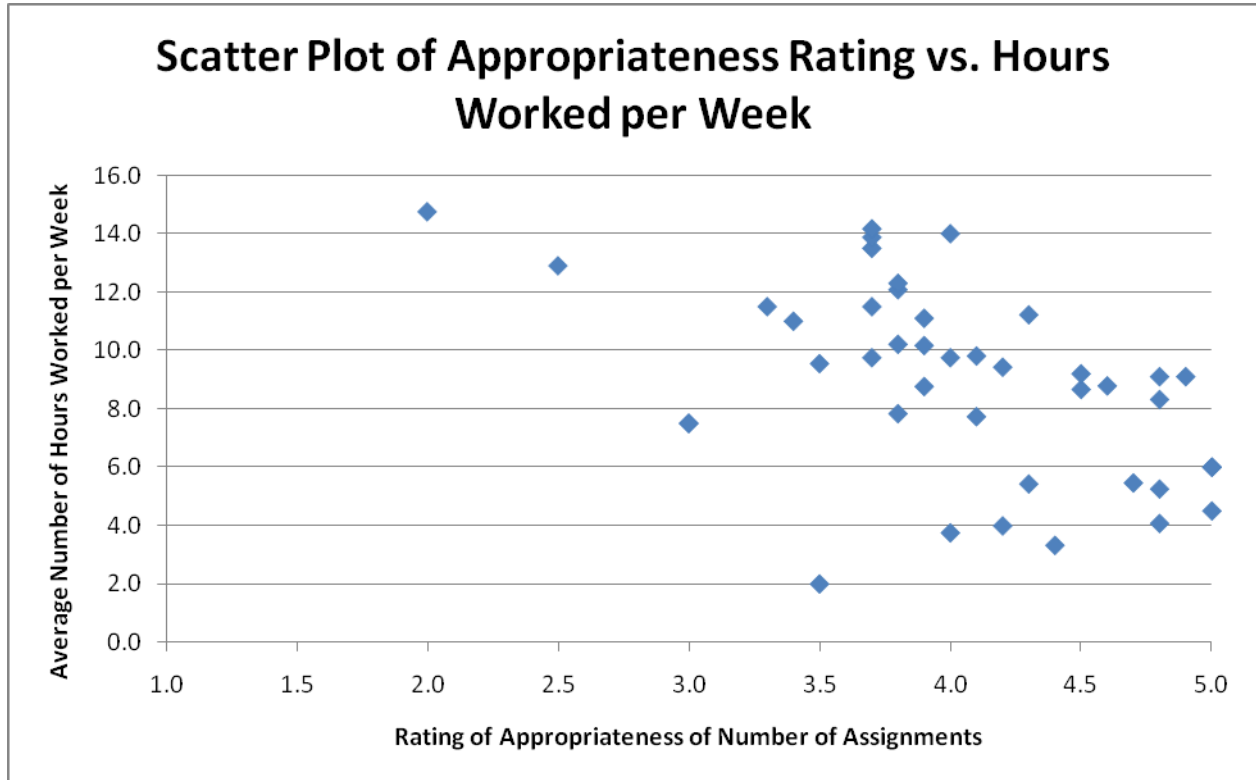
Similar Item Comparison: Work Load

CIOS item 7:	“The number of course assignments (or projects or papers) was appropriate.”
DM PILOT Item Student Effort 1:	“On average, how many hours did you spend on this course per week (total per week in class, on homework, etc.)?”

Data Analysis:

- Because one of these items generated a number in hours rather than a rating scale response, interpolated medians could not be compared. Instead, a correlation between average hours spent on the course per week and ratings of appropriateness of number of course assignments was calculated. This calculation yielded a moderate, negative correlation of $r = -.50$, suggesting that the relationship between number of hours worked and ratings of appropriateness was in the negative direction, such that students tended to provide lower appropriateness ratings as the number of hours worked per week increased.
- Please see Figure 2 for a plot of these data.

Figure 2. Work Load Item Comparison.

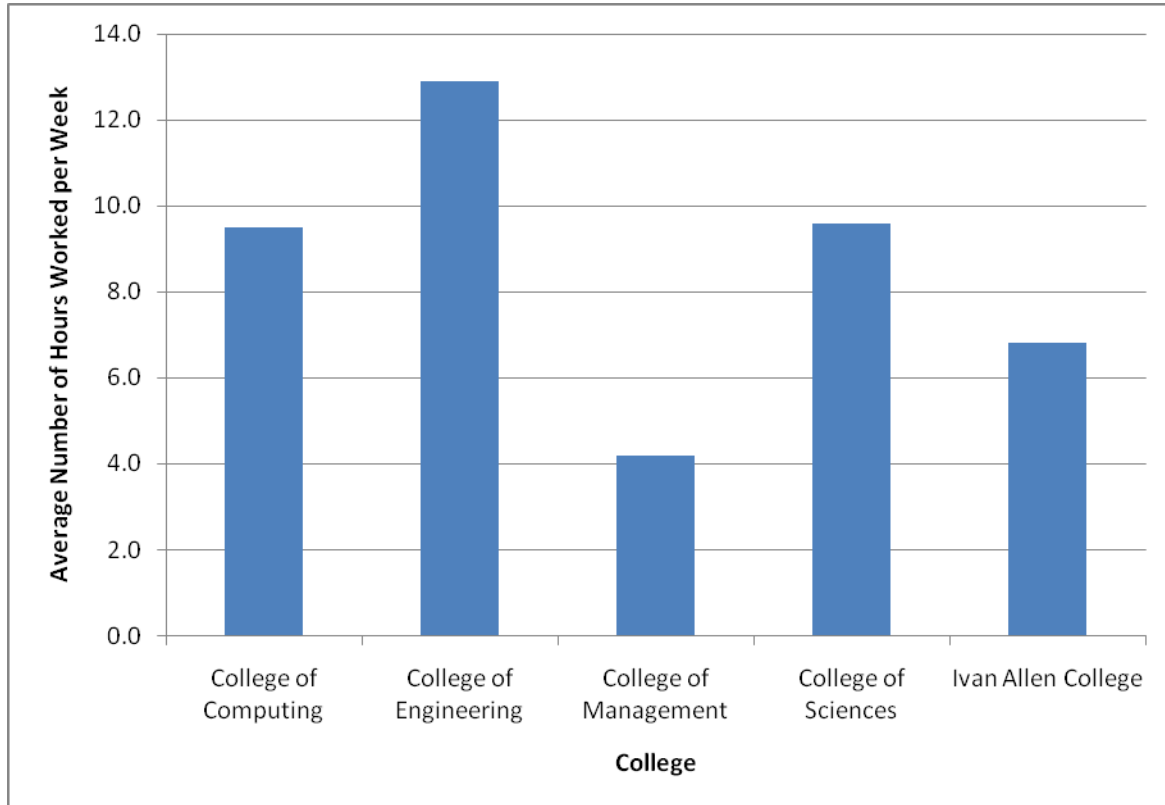


- Correlations were calculated in an effort to see how estimates of workloads for the two surveys related to ratings of overall instructor effectiveness:
 - For CIOS, the correlation between responses to the workload item (“The number of course assignments (or projects or papers) was appropriate”) and Item 10 (“Considering everything, the instructor was an effective teacher”) was moderately strong, $r = .59$.
 - For the DM PILOT, the correlation between responses to the workload item (“On average, how many hours did you spend on this course per week (total per week in class, on homework, etc.)?”) and Item 10 (“Considering everything, the instructor was an effective teacher”) was fairly small, $r = -.11$.
 - This analysis indicates that the student opinions about overall effectiveness and appropriateness of workload are strongly related, whereas opinions about overall effectiveness and actual estimated workload (count of hours) are much less related.
- Workload as measured in the DM PILOT (# hours worked per week) was averaged separately for each of the colleges (please see Figure 3 for a visual representation of these data):
 - The highest workload was reported by students in the College of Engineering, with an average response of 12.9 hours per week (based on data from 129 students in 6 sections).
 - The next highest workload was reported by students in the College of Sciences with an average response of 9.6 hours per week for (based on data from 182 students in 4 sections). An almost equivalent workload was reported by students in the College of

Computing, who worked an average of 9.5 hours per week (based on data from 355 students in 21 sections).

- Students in the Ivan Allen College of Liberal Arts reported an average workload of 6.8 hours per week (based on data from 34 students in 4 sections)
- Students in the College of Management reported the lowest workload, with an average of 4.2 hours per week (based on data from 100 students in 3 sections).

Figure 3. Work Load By College.



- Three of these five figures, College of Computing, College of Sciences, and Ivan Allen College, support the common adage of spending 2-3 hours out of class for each hour in class in order to do well in classes at the university level. Assuming that most of these courses are 3.0 hour courses, and ignoring any lab or recitation requirements, the average workloads reported for these three colleges (9.5, 9.6, and 6.8) correspond roughly to the range you would get from multiplying the number of hours in class by 2-3 hours per week outside of class (2-3 times 3.0 equals 6 to 9 hours per week).
- The College of Engineering average is substantially higher than the 6-9 hours a week range, with students from this college reporting working an average of 12.9 hour per week. This could be a reflection of a high workload and level of difficulty present in the engineering courses included in this data set.

- The College Management average is slightly lower than the 6-9 hours a week range, with students from this college reporting working an average of 4.2 hour per week.

Similar Item Comparison: Relevance

CIOS item 8:	“The examinations and quizzes (or other evaluations) covered the course content and objectives.”
DM PILOT item Quality of Course 3:	“Degree to which activities and assignments facilitated learning.”

Data Analysis:

- Results on these two items were slightly inconsistent for a sizeable portion of the course sections. The majority of courses for which a comparison was possible had a difference of 0.2 or smaller (22 out of 39, or 56%), but the percentage was lower than was seen with the previous item comparisons.
- Of the 39 classes for which a direct comparison was possible, the categorization of course sections based on size differences in the interpolated medians for the two items is as follows:

Item Comparison For Relevance		
Difference Size Category:	Count:	Percentage:
No Difference:	9	23%
Difference = 0.1 - 0.2:	13	33%
Difference = 0.3 - 0.5:	11	28%
Difference = 0.5 - 1.0:	6	15%
Difference > 1.0:	0	0%

- The largest difference between interpolated medians on these items for any section was .9.
- This set of items has lower consistency than the previous sets of items. A likely explanation for this difference is that this set of items is less equivalent than the previous sets. One of these items pertains to exams and quizzes while the other deals with assignments and activities. It seems reasonable that some students may feel quite differently about the assignments and the evaluations they were given in a course. Furthermore, one item asks about covering course objectives and contents while the other asks about facilitating learning. Although a strong relationship between these two ideas should be expected, it is possible that students conceptualized these two evaluative components differently.
- There was a trend towards higher scores on the CIOS item as compared to the DM PILOT item; for the 30 classes in which responses differed between the two surveys, 10 (33%) showed a higher interpolated median response on the DM PILOT item and 20 (66%) showed a higher interpolated median response on the CIOS item. This finding lends further support to the idea that students interpreted these two items differently.

Similar Item Comparison: Content Level

CIOS item 9:	“The examinations and quizzes (or other evaluations) were of appropriate difficulty.”
DM PILOT item Quality of Course 4:	“Degree to which exams, quizzes, homework (or other evaluated assignments) measured your knowledge and understanding”

Data Analysis:

- Results on these two items were slightly inconsistent for a sizeable portion of the course sections. The majority of courses for which a comparison was possible had a difference of 0.2 or smaller (20 out of 39, or 51%), but the percentage was lower than was seen with the previous item comparisons. As was the case with the Relevance items, the larger discrepancy between the items seems reasonable given that the items differ both in terms of what the students are evaluating and on what basis the evaluation is being done. The CIOS item refers specifically to exams and quizzes being of appropriate difficulty, while the DM PILOT item is more general, both in terms of the types of course activities it includes (i.e., exams, quizzes, homework, and other evaluated assignments), and the basis of the evaluation (i.e., appropriateness of difficulty on the CIOS item vs. measurement of knowledge & understanding on the DM PILOT item).
- Of the 39 classes for which a direct comparison was possible, the categorization of course sections based on size differences in the interpolated medians for the two items is as follows:

Item Comparison For Content Level		
Difference Size Category:	Count:	Percentage:
No Difference:	7	18%
Difference = 0.1 - 0.2:	13	33%
Difference = 0.3 - 0.5:	15	38%
Difference = 0.5 - 1.0:	3	8%
Difference > 1.0:	1	3%

- The largest difference between interpolated medians on these two items for any section was 1.5. This is a larger difference than was seen with any other set of items and seems to represent an outlier. Aside from this one large difference, the 2nd largest was 0.6, which is more in line with what was seen with the other sets of items.
- There was no clear trend toward higher responses on this item in one survey or the other; for the 32 classes in which responses differed between the two surveys, 15 (47%) showed a higher interpolated median response on the DM PILOT item and 17 (53%) showed a higher interpolated median response on the CIOS item.

Comparison of Student Comments in CIOS and DM PILOT

The comments from both surveys were analyzed for a representative sample of 12 classes. Comments were designated as “actionable,” meaning that they contained sufficiently specific information to guide the instructor as to what to do (either continue the same or do something different) in order to address the comment, or “not actionable,” meaning they were too imprecise or off-topic to provide any concrete information as to what the instructor should do. The number of comments per survey respondent, number of actionable comments per survey respondent, and % of comments provided which are actionable were calculated for the set of comments on CIOS and the DM PILOT for each of the classes in this set. These values are provided in Table 3.

This analysis reveals that students provided substantially more comments on the DM PILOT than on the CIOS, even though the DM PILOT has fewer actual comment boxes than does the CIOS. We believe that this is driven by two factors: 1) the 200-character limit on most of the CIOS comments was lifted; there are no character limits on any of the DM PILOT items, and 2) the DM PILOT provides specific guidance for each comment box (e.g., “What was the greatest strength,” “What is the most needed improvement,” and requests for specific comments about quality of teaching and quality of course). These two factors seem to have worked together to yield a much higher number of responses per survey on the DM PILOT as compared to the CIOS. The DM PILOT also yielded a larger number of actionable comments per survey, suggesting that DM PILOT data will provide more specific information to instructors about how to both maintain good elements of their teaching and improve their teaching.

Calculations of the percentage of total comments that were actionable revealed that this figure was higher for CIOS. So among the larger set of comments yielded by the DM PILOT, a smaller percentage of them are actionable than was the case with CIOS. However, even a comment which is not be actionable might still be useful to an instructor; for example, comments like “great teacher” are not considered actionable, but are certainly nice to hear! Overall, the body of comments generated by the DM PILOT seems preferable to that generated in CIOS.

Table 3. Comment analysis for CIOS and DM PILOT.

Course:	Digital Measures			CIOS		
	average # comments per survey	average # actionable comments per survey	% actionable comments	average # comments per survey	average # actionable comments per survey	% actionable comments
1 (N = 28)	3.9	1.4	35.3%	0.8	0.4	50.0%
2 (N = 16)	4.8	2.3	48.1%	1.4	0.9	63.6%
3 (N = 85)	3.4	1.6	45.2%	1.5	0.8	56.7%
4 (N = 12)	4.7	2.4	51.8%	2.3	1.9	81.0%
5 (N = 48)	3.8	2.0	52.9%	1.3	1.1	78.4%
6 (N = 18)	3.5	1.7	48.6%	1.1	0.6	50.0%
7 (N = 35)	4.6	3.1	67.2%	3.5	1.9	54.8%

8 (N = 23)	2.8	1.6	56.8%		0.3	0.3	100.0%
9 (N = 56)	3.2	1.6	48.9%		0.4	0.3	83.3%
10 (N = 30)	3.5	1.6	46.6%		0.6	0.3	42.9%
11 (N = 99)	3.1	1.4	46.5%		1.1	0.5	45.5%
12 (N = 4)	4.3	3.5	82.4%		0.3	0.3	100.0%

Appendix A: CIOS and DM PILOT Survey Items

CIOS Core Questions	Scale
1. The course seemed well planned and organized. Comments:	strongly disagree ... strongly agree
2. The instructor did a good job of covering the course objectives and content. Comments:	strongly disagree ... strongly agree
3. The instructor explained complex material clearly. Comments:	strongly disagree ... strongly agree
4. The instructor was approachable and willing to assist individual students. Comments:	strongly disagree ... strongly agree
5. The instructor encouraged students to consult with him or her. Comments:	strongly disagree ... strongly agree
6. Class attendance was important in promoting learning of the material in this course. Comments:	strongly disagree ... strongly agree
7. The number of course assignments (or projects or papers) was appropriate. [Include comments below if you disagree - were there too many or too few for example.] Comments:	strongly disagree ... strongly agree
8. The examinations and quizzes (or other evaluations) covered the course content and objectives. Comments:	strongly disagree ... strongly agree
9. The examinations and quizzes (or other evaluations) were of appropriate difficulty. [Include comments below if you disagree - were they too easy or too hard, for example.] Comments:	strongly disagree ... strongly agree
10. Considering everything, the instructor was an effective teacher. Comments:	strongly disagree ... strongly agree
Comments about the class, instructor, or other issues.	

DM PILOT Items	Scale
<p><u>Student Effort</u></p> <p>1. On average, how many hours did you spend on this course per week (total in class, on homework, etc.)?</p> <p>2. What percentage of classes did you attend?</p> <p>3. What percentage of the homework did you complete?</p> <p>4. Comments about your responses in this section: (eg - were expected and expended effort appropriate for this course?) (there is space for other overall comments later)</p>	<p>0-3, 3-6, 6-9...</p> <p>0-30, 30-50, 50-70, 70-80, 80-90, 90-100</p> <p>0-30, 30-50, 50-70, 70-80, 80-90, 90-100</p> <p>open-ended</p>
<p><u>Quality of Teaching</u></p> <p>1. Instructor's clarity in discussing or presenting course material:</p> <p>2. The instructor clearly communicated what it would take to succeed in this course.</p> <p>3. Instructor's respect and concern for students:</p> <p>4. Instructor's level of enthusiasm about teaching the course:</p> <p>5. Instructor's ability to stimulate my interest in the subject matter.</p> <p>6. The instructor was readily available for consultation.</p> <p>7. Helpfulness of feedback on assignments.</p> <p>8. Considering everything, the instructor was an effective teacher.</p> <p>9. What was the greatest strength?</p> <p>10. What is the most needed improvement?</p> <p>8. Comments about your responses in this section (quality of teaching): (there is space for other overall comments later)</p>	<p>very poor ... exceptional</p> <p>strongly disagree ...strongly agree</p> <p>very poor ... exceptional</p> <p>detached ... extremely enthusiastic</p> <p>ruined my interest ... made me eager to learn more</p> <p>strongly disagree ...strongly agree</p> <p>not helpful ... extremely helpful</p> <p>strongly disagree.....strongly agree</p> <p>open-ended</p> <p>open-ended</p> <p>open-ended</p>
<p><u>Quality of Course</u></p> <p>1. Rate how prepared you were to take this subject:</p> <p>2. How much would you say you learned in this course?</p> <p>3. Degree to which activities and assignments facilitated learning:</p> <p>4. Degree to which exams, quizzes, homework (or other evaluated assignments) measured your knowledge and understanding:</p> <p>5. Considering everything, this was an effective course.</p> <p>6. What was the best aspect?</p> <p>7. How could it be improved?</p> <p>8. Other comments about your specific responses in this section (quality of course): (there is space for other overall comments later)</p>	<p>completely unprepared ... extremely well prepared</p> <p>almost nothing ... an exceptional amount</p> <p>very poor ... exceptional</p> <p>very poor ... exceptional</p> <p>useless ... essential</p> <p>strongly disagree...strongly agree</p> <p>open-ended</p> <p>open-ended</p> <p>open-ended</p>
<p><u>Overall Comments</u></p> <p>1. Other overall comments:</p>	<p>Open-ended</p>