CEPP Report on Quantitative Student Evaluation of Teaching (qSET) Analysis

April 2019

Table of Contents

Brief Background

The College has long sought student feedback to evaluate teaching at Skidmore. The College, relevant committees, and faculty adopted a 3-question format over two decades ago to gain insight into students' evaluation of teaching effectiveness. In an effort to assist faculty in receiving more nuanced, and less biased, feedback CEPP and ATC (formerly known as CAPT) formed a joint subcommittee in 2011 to evaluate and potentially revise the 3-question student evaluation of teaching instrument. Based on the work of this subcommittee, along with CEPP input, internal and external consultation, and pilot testing, the current "Student Rating of Courses and Teaching" form was born and ultimately adopted by the faculty (See: https://www.skidmore.edu/dof-vpaa/meetings/faculty/2012-2013/minutes3-1.php). The rationale outlined in the motion put forth by CEPP proposed that "an assessment of this new form be conducted no later than the academic year 2016/2017, with that assessment shared with the faculty." Likely due to CEPP's focus on the general education curriculum overhaul, this assessment was postponed until the current academic year. To conduct this analysis of the quantitative student evaluation of teaching (qSET), CEPP devised the following process outlined in the figure below:

Historical Study

To understand the genesis of the current qSET form, CEPP read the historical context and the work that went into developing and testing the form. Michael Arnush, the faculty member who chaired CEPP during the development of the current qSET, met with CEPP in spring 2018. Subsequently, a representative from CEPP (S.Ives, Co-Chair) also met with Catherine Berheide, a faculty member who was directly involved in the refinement and testing of the instrument. In summary, the current qSET was piloted in the 2012 FYE in conjunction with a National Science Foundation award, and was developed and refined in broad consultation with faculty internally, and through seeking input from external experts in the field as well (Joey Sprague and Catherine Ross). In the spirit of following a similar process, we sought to perform an internal analysis of the qSET form

Internal Assessment of qSET

There a very few topics that bring about such passionate discourse as the topic of student evaluations of teaching, so it is no surprise the topic has been a longstanding and active area of scholarly investigation. Published work has raised concerns about the following: what such forms are actually measuring [1], potential bias against women [2], potential bias against people of color [3], potential bias against non-native English speakers [4], disparity between disciplines, and potential issues in statistical interpretation [5]. Thus, CEPP felt it prudent to conduct an internal analysis of the qSET data over the last 5 years (2013-2018) since the form was adopted.

Analysis of qSET Data (2013-2018)

CEPP partnered with the Office of Institutional Research to gather and analyze the qSET data, as well as data on potential factors that might contribute to or explain some of the scores (e.g. academic division). The specific information gathered and the methods used to analyze the

- Average qSET scores are generally quite high (~4.5 out of 5), with the bulk of individual ratings falling in the 4-5 range, but the scores still vary by nearly a full point when considering student-level data. The variation in scores within a question is an important consideration, as it is unlikely that small deviations in qSET scores represent significant deviations from the average.
- qSET scores trend upward over time, increasing about 3% from 2013 to 2018.
- Consistent with the pilot study of the current form, there is a very high degree of congruity (correlation ~0.94) between a composite score of sub-questions and the independent "overall" questions for "the course," for example.

In terms of student, course, and instructor characteristics, several factors were found to be statistically significant (i.e. not likely due to chance alone), but the magnitude of the effects were relatively small compared to the overall variation in scores (see table appended below). The items that had the largest effect were "students desire to enroll in the course" and "expected grade," both of which were positively related to qSET scores, meaning the greater the desire to enroll in the course and the higher the expected grade, the higher the qSET scores. The increase in qSET scores was about 0.1 and 0.3 on a 5-unit scale for each incremental change in "expected grade" or "desire to enroll in the course," respectively. To put this into context, if students in a class all reported one unit greater desire to enroll in the course—e.g. "more than other courses" (a 4 on the qSET form) vs. "about the same" (a 3) [illegible]

being explored (e.g. does time of day influence qSET scores). Specifically, based upon probability, the more questions we ask or variables we explore, the more likely we might find a "significant" result simply because we asked many questions or explored many variables. In statistics, many would suggest that researchers adjust their level of significance, or alpha level, to compensate for the number of questions or variables being probed [6]. Such an adjustment to what was deemed statistically significant, either a priori or post hoc, was not made in the current analysis.

The level of the course does appear to influence qSET scores, since scores appear to fall as the course level increases. 200- and 300-level courses are significantly scored lower than 100-level courses.

There are some additional factors that were looked at. For example, instructor status (e.g. instructor, tenure-track, or tenured) was never significant. Another item for consideration was expected grade vs. actual grade and the effects and trends were similar. Interestingly, neither expected nor actual grades differed over time.

Though most effects were a fraction of a point and some were statistically insignificant, it may still be that an individual instructor's scores are impacted by the sum of the various factors. For example, an older professor teaching a required course (low desire to enroll) at an upper level may see lower scores as a result of the combined effect of these course and instructor characteristics.

As is often the case with research, a project can bring about more questions than it answers, and the robust analysis conducted in collaboration with the Office of Institutional Research is not meant to either exonerate or condemn the qSET necessarily, but certainly provides information deserving of consideration by key parties (ATC, PC, etc.) and the faculty as a whole. The outputs of the analysis will be made fully available to the Skidmore community (https://www.skidmore.edu/ir/course-evaluations/), though to protect confidentiality of individual instructors scores, the raw data will not be made publicly available.

Analysis of Faculty Knowledge and Perceptions of the qSET form

In accordance with guidance set forth in the CEPP operating code, which states that CEPP will "ensure extensive, widespread and high quality consultation take place during all major initiatives," we devised and conducted a survey of all faculty to gain insight into faculty perceptions of the Skidmore qSET and qSETs more broadly. This mechanism was designed to be more inclusive and anonymous in nature, and gain feedback from a larger number of faculty than typically engage in faculty floor or open forum discussions.

The survey was made available to the faculty email list. From this, 168 faculty completed the survey, with 48 non-tenure-track (TT), 39 tenure-track, and 80 tenured faculty. Approximately 40% were current or former department chair/program directors and about 15% had served on the tenure and promotion committee. Collectively, the sample was relatively robust and representative of the College.

References

1. Uttl, B., C.A. White, and D.W. Gonzalez, *Meta-*

Table Summarizing 2013-2018 qSET Data Analysis

	Course Overall			Instructor Overall			Learning Overall		
Independent Variable	Effect on Score?	# of years significant	Effect (%)	Effect on Score?	# of years significant	Effect (%)	Effect on Score?	# of years significant	Effect (%)
Academic Division (relative to Humanities)									
Physical & Life Sciences	-0.05	3	-1.1	-0.06	4	-1.4	-0.03	1	-0.7
Social Sciences	-0.01	0	-0.3	-0.03	1	-0.6	-0.01	0	-0.1
Visual & Performing Arts	-0.05	3	-1.0	-0.06	4	-1.4	-0.04	2	-0.9
Other	-0.06	5	-1.3	-0.07	4	-1.5	-0.04	1	-1.0
Course Level (relative to 100-Level)									
200-Level Courses	-0.04	3	-0.8	-0.03	1	-0.6	-0.03	2	-0.7
300-Level Courses	-0.06	5	-1.4	-0.04	1	-0.8	-0.07	5	-1.5
Start Time (relative to Earliest morning <9am)									
Early morning (9am-9:45am)	0.00	0	0.0	0.00	1	0.1	0.00	0	0.0
Late morning (10am-11:55am)	0.01	0	0.1	0.00	0	0.1	0.01	0	0.2
Early afternoon (12pm-2:50pm)	-0.01	1	-0.2	0.00	1	0.0	0.00	1	-0.1
Late afternoon (3pm-4:40pm)	-0.01	0	-0.2	-0.01	0	-0.2	0.00	0	-0.1
Evening (5pm-8:30pm)	-0.01	1	-0.2	-0.01	2	-0.2	-0.01	1	-0.2
Reason Taking Course									
Required for major	0.00	0	0.0	0.00	0	0.0	0.00	0	0.0
Elective for major	0.00	0	0.0	0.00	0	0.0	0.00	0	0.0
Non-major elective	0.00	0	0.0	0.00	0	0.0	0.00	0	0.0
All-college requirement	0.00	1	0.0	0.00	2	0.0	0.00	1	0.0
Other requirement	0.00	0	0.0	0.00	0	0.0	0.00	0	0.0
Other Course Characteristics									
Course Enrollment #	0.00	4	-0.1	0.00	2	-0.1	0.00	5	-0.1
Course Credits #	-0.02	2	-0.5	-0.03	1	-0.7	-0.02	2	-0.3
Fall Course	-0.01	1	-0.2	-0.01	0	-0.1	-0.01	2	-0.2
Student Information									
Desire to enroll in course	0.30	5	6.7	0.25	5	5.5	0.33	5	7.5
Hours/week devoted to course	0.02	5	0.5	0.01	2	0.3	0.03	5	0.8
Expected grade (0.00-4.00)	0.20	5	4.4	0.20	5	4.4	0.15	4	3.3
Course Composition									
Class Year (1-4 avg)	-0.01	0	-0.2	-0.01	0	-0.2	0.01	1	0.3
% Class ALANA	0.00	1	0.0	0.00	1	0.0	0.00	2	0.0
% Class Female	0.00	3	0.0	0.00	3	0.0	0.00	3	0.0
Response Rate	0.00	2	0.0	0.00	2	0.0	0.00	2	0.0
Instructor									
Instructor ALANA	-0.05	3	-1.1	-0.06	3	-1.3	-0.04	3	-1.0
Instructor International	-0.03	1	-0.6	-0.03	0	-0.7	-0.02	0	-0.4
Instructor Female	0.01	0	0.2	0.01	0	0.3	0.01	0	0.2
Instructor Age	0.00	5	-0.1	-0.01	5	-0.1	0.00	4	-0.1
Employment Category (relative to Fulltime-Regular)									
Fulltime-Temporary	-0.04	2	-0.8	-0.04	2	-1.0	-0.04	3	-0.8
Parttime-Regular or Shared Parttime-Regular	0.01	0	0.3	0.01	0	0.2	0.02	0	0.4
Parttime-Temporary	-0.09	5	-2.1	-0.10	5	-2.3	-0.08	5	-1.8

Introduction to Appendices

In support of the report released by CEPP via email (April 23, 2019) and presented to the faculty (April 26, 2019), the following appendices provide a more complete picture of the findings that CEPP reviewed before issuing the report. Not everyone is familiar with these types of data analyses, and this introduction to the appendices provides some context. All analyses are at the level of the course section, not at the level of the student. So, instead of roughly 2,500 students taking ~8 courses per year, totaling ~20,000+ student responses/year, there are ~1,300-1,400 data points per academic year reflecting the number of independent courses sections. There are several reasons for this, one of which is the ability to capture the demographic composition of each course section in the analysis. All analyses are at the course section level with the exception of Appendix B which is at the student level and matches the all-college means on individual faculty course evaluation results and all-college reports at: https://www.skidmore.edu/ir/course-evaluations.

A question that CEPP had in the course of its assessment was whether the specific sub items, such as "The course….was well organized" etc., related to the independent "course overall" question. To do this composite scores, or variables, of the sub items were cre (.e)4 (du/)e (o)- independent " 4 (P)-o4 (

Finally, as outlined in the report, CEPP was curious whether course factors (e.g., enrollment number, level, etc.), student factors (e.g., expected grade, etc.), or instructor characteristics (e.g., gender, race, etc.) influence qSET scores. The analytical approach was a multivariate regression analysis, which essentially gauges how much a given factor might explain or influence the qSET scores, or more specifically the variation in the qSET scores, while controlling for other factors in the model. Similar to correlation, regression coefficients range from 0 to 1, 0 would mean a factor explains or influences 0% of the qSET scores; whereas, a coefficient of 1 would indicate that the factor explains 100% of the qSET scores. This analysis was performed for each academic year (2013-14 to 2017-

Appendix A: Bivariate Correlations

Bivariate Correlations – 2017-18

All years (2013-14 to 2017-18) are in the 0.93-0.95 range. 2017-18 presented to illustrate.

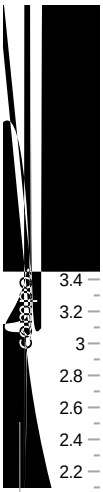
Bivariate Fit of composite_var_course By Q02.01_mean academic_year=17-18

Course - Mean of Q01.01_mean thru Q01.09_mean

Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
Q02.01_mean	4.477825	0.42837	0.938958	<.0001*	1373
composite_var_course	4.420019	0.370416			

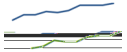
Bivariate Fit of composite_var_instr By Q04.01_mean academic_year=17-18

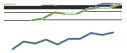
Instructor - Mean of Q03.01_mean thru Q03.10_mean

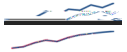


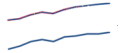
Variable

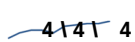
Appendix B: Trends in Course Evaluation Means

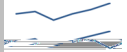


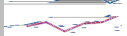












All-College (per ClassClimate)
Means

		Fall 2010	Spring 2011	Fall 2011	Spring 2012	Fall 2012	Spring 2013	Fall 2013	Spring 2014	Fall 2014	Spring 2015	Fall 2015	Spring 2016	Fall 2016	Spring 2017	Fall 2017	Spring 2018
1. The Course...																	
1.01 content was well organized.	Mean							4.26	4.31	4.34	4.34	4.33	4.37	4.37	4.38	4.39	4.43
1.02 objectives were met.	Mean							4.33	4.37	4.40	4.41	4.39	4.43	4.44	4.45	4.45	4.50
1.03 materials (e.g., readings, handouts, videos) contributed to my learning.	Mean							4.29	4.32	4.35	4.36	4.35	4.38	4.39	4.42	4.43	4.43
1.04 helped me learn concepts and methods.	Mean							4.33	4.36	4.38	4.39	4.38	4.42	4.43	4.45	4.45	4.49
1.05 improved my ability to communicate clearly about the subject.	Mean							4.26	4.33	4.34	4.37	4.35	4.41	4.40	4.43	4.42	4.47
1.06 enabled me to think independently about the subject matter.	Mean							4.27	4.34	4.34	4.37	4.36	4.40	4.41	4.44	4.44	4.47
1.07 assignments helped me achieve the course objectives.	Mean							4.23	4.29	4.31	4.33	4.32	4.35	4.37	4.39	4.39	4.42
1.08 q ş ş ş																	
	Mean							4.52	4.54	4.58	4.57	4.55	4.58	4.59	4.61	4.61	4.63
3.03 answered questions effectively.	Mean							4.39	4.44	4.43	4.46	4.42	4.47	4.48	4.50	4.49	4.52
3.04 stimulated interest in the subject.	Mean							4.36	4.41	4.41	4.44	4.43	4.45	4.50	4.50	4.50	4.52
3.05 gave assignments related to the course objectives.	Mean							4.42	4.43	4.47	4.46	4.46	4.49	4.50	4.50	4.53	4.53
3.06 provided clear criteria for grading.	Mean							4.12	4.20	4.18	4.22	4.17	4.23	4.23	4.29	4.27	4.30
3.07 was available outside of class (e.g., office hours, by appointment, email).	Mean							4.43	4.43	4.46	4.47	4.48	4.50	4.49	4.53	4.54	4.55
3.08 suggested ways students could improve.	Mean							4.25	4.30	4.31	4.35	4.31	4.37	4.36	4.42	4.39	4.44
3.09 fostered an environment of respect in the classroom.	Mean							4.46	4.47	4.50	4.52	4.51	4.54	4.56	4.57	4.58	4.59
3.10 set high standards for students.	Mean							4.37	4.40	4.45	4.47	4.45	4.50	4.51	4.53	4.53	4.55
Q1 Overall, I learned...																	
	Mean	4.36	4.37	4.33	4.36	4.38	4.41										
Q2 Overall, the teaching was...	Mean	4.22	4.24	4.18	4.23	4.26	4.29										
Q3 Overall, the course was...	Mean	4.12	4.15	4.08	4.14	4.15	4.21										

Evals Fall 2013 onward Scale: Disagree Strongly to Agree Strongly

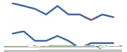
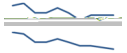
Evals Prior to Fall 2013 Scales:

Learned = A great deal, quite a bit, some, not very much, very little

Teaching & Course = Excellent, Very Good, Good, Fair, Poor

All-College (per ClassClimate)

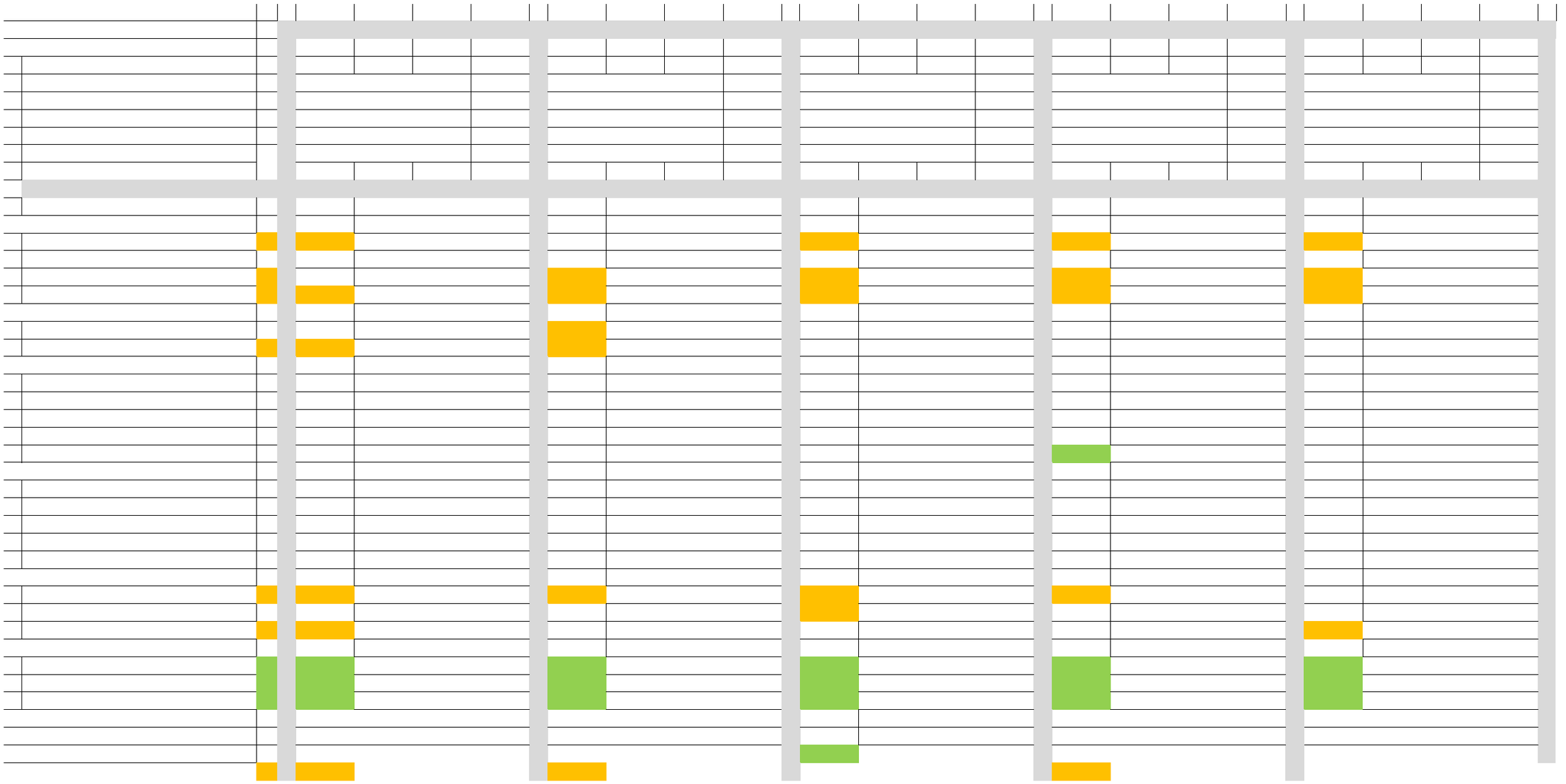
Standard Deviations

		Fall 2010	Spring 2011	Fall 2011	Spring 2012	Fall 2012	Spring 2013	Fall 2013	Spring 2014	Fall 2014	Spring 2015	Fall 2015	Spring 2016	Fall 2016	Spring 2017	Fall 2017	Spring 2018	
1. The Course...																		
1.01 content was well organized.	S.D.							0.83	0.80	0.78	0.80	0.82	0.78	0.79	0.79	0.79	0.77	
1.02 objectives were met.	S.D.							0.74	0.73	0.70	0.71	0.73	0.71	0.70	0.71	0.70	0.68	
1.03 materials (e.g., readings, handouts, videos) contributed to my learning.	S.D.							0.84	0.82	0.80	0.79	0.81	0.80	0.79	0.78	0.78	0.80	
1.04 helped me learn concepts and methods.	S.D.							0.79	0.78	0.77	0.77	0.77	0.76	0.76	0.75	0.75	0.75	
1.05 improved my ability to communicate clearly about the subject.	S.D.							0.83	0.80	0.79	0.79	0.80	0.78	0.78	0.77	0.78	0.75	
1.06 enabled me to think independently about the subject matter.	S.D.							0.82	0.80	0.80	0.79	0.79	0.77	0.76	0.76	0.76	0.75	
1.07 assignments helped me achieve the course objectives.	S.D.							0.84	0.82	0.79	0.80	0.82	0.80	0.79	0.79	0.79	0.79	
1.08 feedback I received helped me achieve the course objectives.	S.D.							0.94	0.90	0.91	0.90	0.92	0.88	0.88	0.86	0.87	0.88	
1.09 was challenging.	S.D.							0.91	0.87	0.85	0.84	0.87	0.83	0.85	0.85	0.84	0.82	
2. Course Overall																		
2.01 Overall, this was an effective course.	**S.D.**							**0.82**	**0.80**	**0.78**	**0.78**	**0.80**	**0.77**	**0.76**	**0.78**	**0.77**	**0.76**	
3. The Instructor...																		
3.01 presented the course material effectively.	S.D.							0.86	0.84	0.82	0.83	0.85	0.82	0.81	0.81	0.81	0.80	
3.02 was prepared for class.	S.D.							0.71	0.68	0.64	0.66	0.70	0.66	0.66	0.64	0.65	0.63	
3.03 answered questions effectively.	S.D.							0.82	0.79	0.81	0.78	0.83	0.79	0.77	0.77	0.77	0.76	
3.04 stimulated interest in the subject.	S.D.							0.87	0.84	0.85	0.82	0.83	0.82	0.78	0.79	0.79	0.80	
3.05 gave assignments related to the course objectives.	S.D.							0.75	0.75	0.71	0.74	0.74	0.72	0.72	0.73	0.70	0.71	
3.06 provided clear criteria for grading.	S.D.							1.01	0.95	0.97	0.95	1.00	0.96	0.96	0.94	0.96	0.96	
3.07 was available outside of class (e.g., office hours, by appointment, email).	S.D.							0.74	0.75	0.73	0.73	0.73	0.73	0.73	0.70	0.69	0.70	
3.08 suggested ways students could improve.	S.D.							0.86	0.85	0.84	0.82	0.85	0.82	0.82	0.80	0.82	0.81	
3.09 fostered an environment of respect in the classroom.	S.D.							0.75	0.76	0.72	0.72	0.74	0.72	0.69	0.71	0.71	0.71	
3.10 set high standards for students.	S.D.							0.80	0.79	0.74	0.74	0.76	0.74	0.72	0.72	0.71	0.70	
4. Instructor Overall																		
4.01 Overall, the instructor was an effective teacher.	**S.D.**							**0.82**	**0.80**	**0.78**	**0.78**	**0.80**	**0.76**	**0.75**	**0.75**	**0.76**	**0.75**	
5. Learning Overall																		
5.01 Overall, I learned																		
																		

Appendix C: Distributions of Dependent Variables

Distributions of Dependent Variables
2017-18

Appendix D: Regression Findings - All

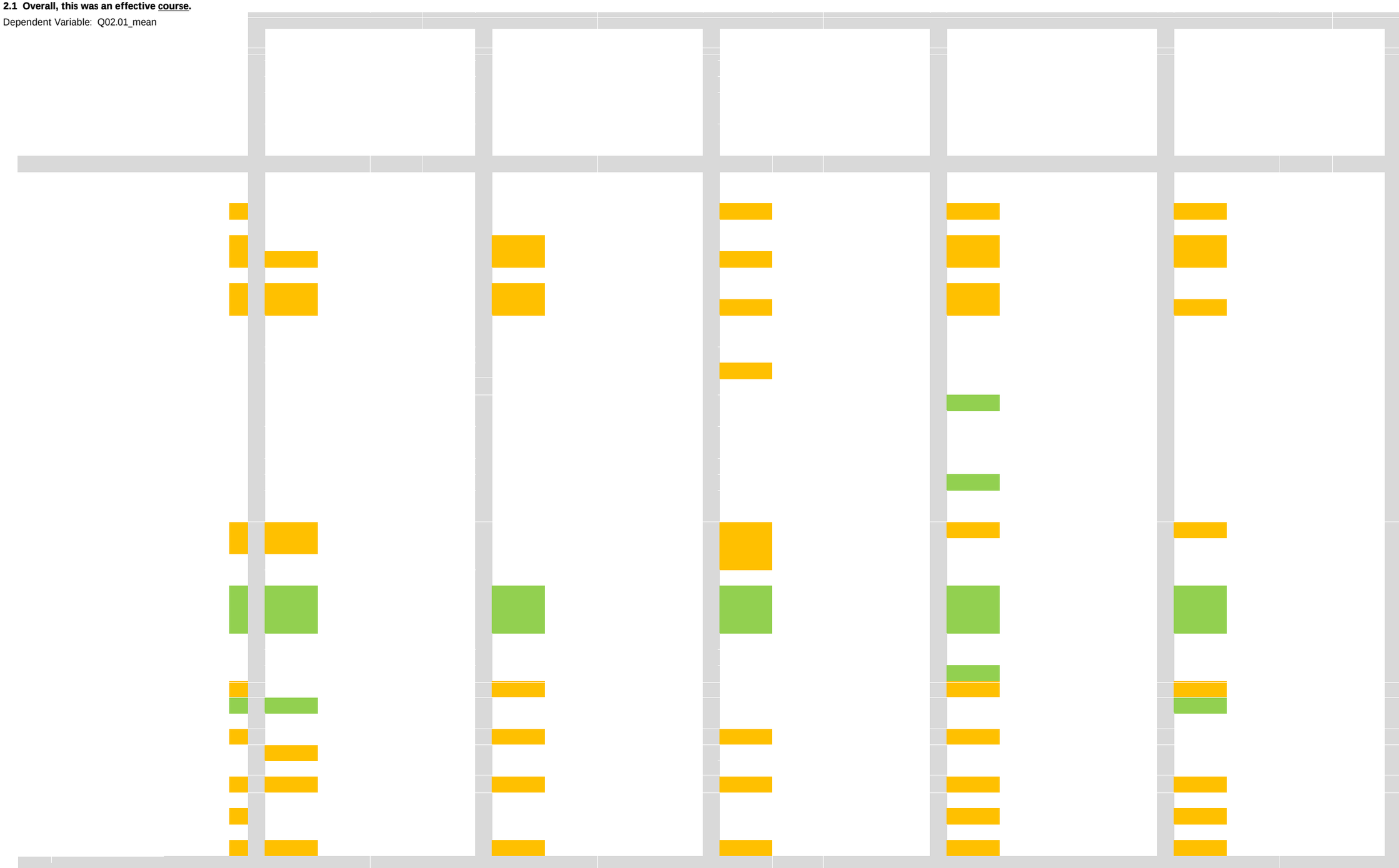


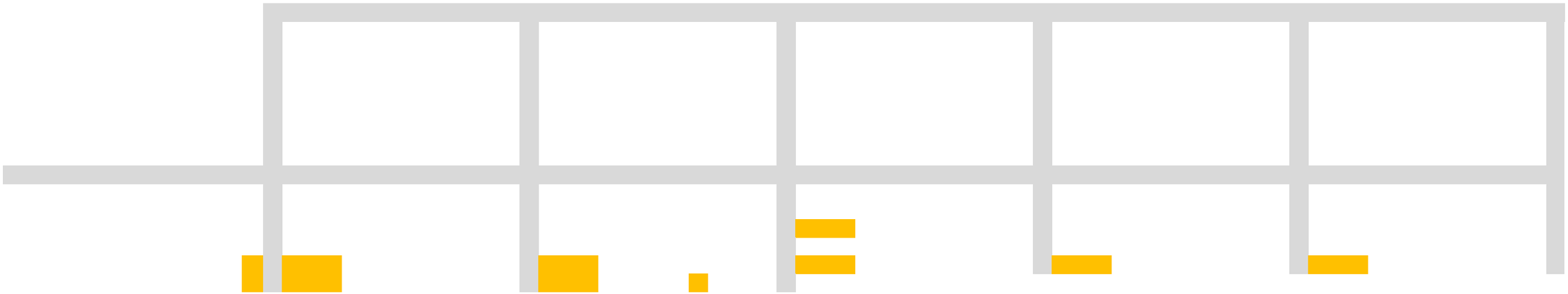
****Composite Course Variable created from:**

Academic Division of 'Other' includes:

2.1 Overall, this was an effective course.

Dependent Variable: Q02.01_mean





4.1 Overall, the instructor was an effective teacher.

Dependent Variable: Q04.01_mean

Random Effect: instructor_id

	2013-14	2014-15	2015-16	2016-17	2017-18
	0.624167	0.671515	0.623679	0.641479	0.620514
	0.614229	0.662919	0.6138	0.632264	0.610828
	0.323139	0.29843	0.315729	0.278284	0.307392
	4.442059	4.475523	4.491609	4.543703	4.525729
	1282	1295	1291	1318	1327

	Estimate	Std Error	t Ratio	Prob>|t|	Estimate	Std Error	t Ratio	Prob>|t|	Estimate	Std Error	t Ratio	Prob>|t|	Estimate	Std Error	t Ratio	Prob>|t|	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.8410027	0.310152	9.16	<.0001*	2.4500152	2	Std		0				E.000		2.450015		2.450015