Summary Teaching Evaluations (Fall 2022)

Summary of Teaching as Instructor of Record

Dalhousie University University of Toronto

	Fall Semester	Winter Semester	Summer Semester
2022-2023	PHIL 2660 **Logic: Understanding Scientific Reasoning** 40† Students	PHIL 2490 / CSCI 3101 **Social, Ethical, and Professional Issues in Computer Science** 15† / 200† Students	
	PHIL 4070 / 5070 **Philosophy of Psychology** (*Topic*: Philosophy on the Spectrum: Philosophy of Autism & Autistic Philosophy) 10† / 5† Students	ASSC / CSCI 1801 **Case Studies in Computing and Society** 400† Students	
2021 - 2022	PHIL 4115 / 5115 **Topics in Ethics I** (*Topic*: Evolutionary Ethics) 4 / 1 Students	PHIL 2490 / CSCI 3101 **Social, Ethical, and Professional Issues in Computer Science** 14 / 213 Students 2 Teaching Assistants	
		ASSC / CSCI 1801 **Case Studies in Computing and Society** 209 Students 7 Teaching Assistants	
2020 - 2021			PHIL 455 **Seminar in the Philosophy of Science** (AI & the Value-Alignment Problem) 12 Students

† Denotes course cap while registration is still in flux.

Total number of students taught over this period: **1123**†

Summary of Teaching as Teaching Assistant

University of California, Irvine Simon Fraser University

	Fall Quarter	Winter Quarter	Spring Quarter	Summer Quarter
2017 - 2018	LING 51 / PSYC 56 **Acquisition of Language** (Lisa Pearl) 152 Students	LING 3 **Intro. to Linguistics** (Arunima Choudhury) 70 Students	LPS 31 / PHIL 31 **Inductive Logic** (Simon Huttegger) 43 Students	
2016 – 2017		ECON 15A **Probability and Statistics for Economics** (Kent Johnson) 75 Students	LPS 31 / PHIL 31 **Inductive Logic** (Simon Huttegger) 44 Students	

	Fall Semester	Spring Semester	Summer Semester
2015 - 2016	PHIL xx1 (now PHIL 105) **Critical Thinking** (Jillian McIntosh) 76 Students	PHIL xx1 (now PHIL 105) **Critical Thinking** (Jillian McIntosh) 63 Students	
2104 - 2015	PHIL 120 **Introduction to Ethics** (Evan Tiffany) 45 Students	PHIL xx1 (now PHIL 105) **Critical Thinking** (Jillian McIntosh) 51 Students	

Total number of students taught over this period: **619**

Evidence of Teaching Effectiveness

In reflecting on my approach to teaching, I attempt to answer the following two questions: *What are the students supposed to learn? How can I accurately measure whether I have been successful in teaching them?* With regard to the first question, I try to create assignments that facilitate training in practical and transferable skills. At the undergraduate level, this includes summarising complex arguments, and focusing in on the key point being argued for and the logical support for that key point. At the graduate level, this may include gaining experience with the practical skills involved in certain academic disciplines—writing papers, to be sure, but also practices that foster collegiality, such as refereeing, providing commentary, presenting work to one's colleagues, etc. As such, the assignments for my courses are designed to reflect these objectives. (See sample syllabi, Appendix A.) I do two things to facilitate learning practical skills in university writing and analysis.

First, I assign short, low-stakes assignments that allow students to gain experience with important skills required for writing university essays.

Second, I either assign *two* of each assignment (for short assignments in lower-level courses), or I break an assignment up into component parts (for longer writing assignments in upper-level courses) so students have multiple opportunities to practise the same skill, while simultaneously learning to take feedback into account.

With this in mind, it is important for me to be able to measure whether these objectives are being satisfied, which means ensuring that the proxies I use in grading actually reflect the true goal of the assignment.

Metrics for Teaching Effectiveness (Qualitative)

Qualitatively, my approach has been successful, as evidenced by the written feedback on the year-end course evaluation. To give a small sample, I quote, unedited, some student responses from my course, ASSC 1801: Case Studies in Computing and Society. This course was taught, virtually, at Dalhousie University in Winter 2022. Housed in the Faculty of Arts and Social Science, this course is required for majors in the Faculty of Computer Science (Bachelor Computer Science and Bachelor Applied Computer Science). For some context, the assignments for this course included two '4-sentence essays' and two 'Tweet assignments' in the first half of the semester, as well as two short essays in the second half of the semester.

For the first assignment, students were given prompts to respond to a specific paper that we had read in class. (See 'sample assignments' in Appendix C.) The idea is to help students understand how to

(1) summarise an author's argument ("they say…"),
(2) respond to that argument with their own opinion and evidence ("I say… Because…"),
(3) consider alternative viewpoints ("One might object…"), and
(4) respond to those counterarguments ("I would respond…").

The Tweet assignment, then, emphasises writing a thesis statement, which requires understanding the key point that an author is making well enough to be able to state it in one's own words with a very strict word limit: 280 characters, including whitespace. (See Appendix C for full example.)

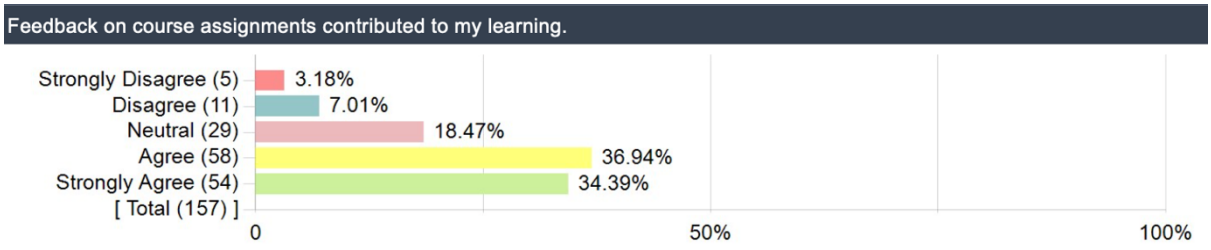
Finally, the two short papers were a standard first-year-paper assignment; however, I ensured that the TAs made explicit in their discussion sections how a 750-word paper can be seen as a mere expansion of the four-sentence essay, and how a thesis statement can be seen as equivalent to the tweets the students composed in the first half of the semester. In general, this appeared to be successful, insofar as some students noticed and appreciated the structure of 'building up' skills over the course of the semester:

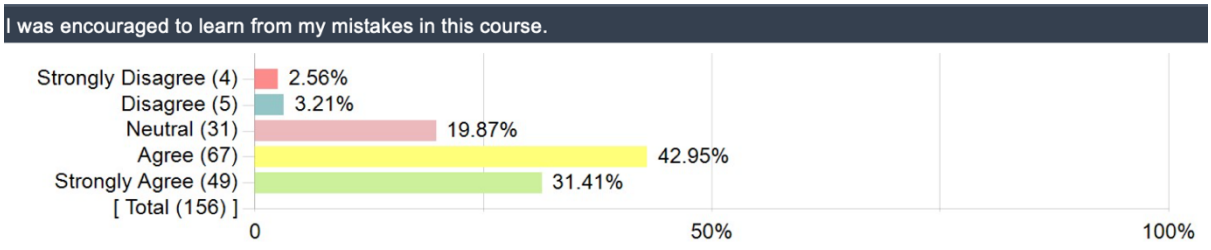
- "I liked the structure of it a lot where we learned how to summarize and structure our argument through the 4–sentence assignments and tweet assignments then moved on to learn how to write an essay"
- "I liked how the tweet assignment and 4 sentence papers built up to the essay"

Several students explicitly said that these short assignments helped them with their writing skills:

- "I felt the four sentence papers were a good exercise for developing writing skills"
- "I think that the Four–sentence paper helps a lot with my essay writing"
- "I liked the structure of [the course] a lot where we learned how to summarize and structure our argument through the 4–sentence assignments and tweet assignments then moved on to learn how to write an essay"
- "The built–in difficulty level of the course is very smooth, from the first 4-sentence-essay to a short paper, I can actually feel I make some progress"
- "The course structure was designed in a way that you were building up knowledge on how to do small tasks (such as constructing themes), up to writing a full essay. I found this structure to be quite helpful"
- "The buildup of smaller assignments to the final short papers really helped me succeed"

These qualitative sentiments were reflected in the quantitative component of the year-end course evaluations, where students were asked whether feedback on assignments contributed to their learning and whether they were encouraged to learn from their mistakes. In both cases, nearly three quarters (0.71; 0.74) of students who responded to this question (n = 157; 156) said that they agreed, or strongly agreed. Only a small fraction (0.10; 0.06) disagreed or strongly disagreed with these statements. This can be seen in the graphs below. (Full course evaluations are provided in Appendix B.)





Metrics for Teaching Effectiveness (Quantitative)

As mentioned above, in my course on Case Studies in Computing and Society (ASSC 1801, Winter 2022), I assigned two of each assignment, so students can get some experience taking feedback into account and retrying the same task. This also provides invaluable data for measuring whether the assignments are actually effective in *teaching* the students practical skills, based upon the evaluation criteria and the skills being measured.

In the proceeding analysis, I will refer to the two four-sentence essays as 'Set 1', the two tweet assignments as 'Set 2', and the two short papers as 'Set 3'.

Considering the students who submitted both assignments in Set 1 ($n_1 = 192$), the students' grades increased, on average, by 6.25%, from an average grade of 'B' on the first 'four-sentence essay' assignment to 'B+' on the second.

For the students who submitted both assignments in Set 2 ($n_2 = 189$), the students' grades increased, on average, by 11.59%, from an average grade of 'C' on the first 'tweet' assignment to 'B' on the second.

Finally, for the students who submitted both assignments in Set 3 ($n_3 = 187$), the students' grades increased, on average, by 2.6%, from an average grade of 'B-' on the first short paper to 'B' on the second.

These increases are suggestive of student learning between the two assignments in each set. However, because this is only an *average* increase, it does not necessarily tell us much about individual performance. We can examine the number of students whose grade increased between the two assignments in each of the sets.

Considering the students who submitted both assignments in Set 1 ($n_1 = 192$), the majority (**0.5260**) did better on the second assignment than the first. If we ignore the 20 students who received a perfect score on the first assignment (because their scores could not increase), then a slightly higher majority (**0.5872**) had their scores increase between the two assignments (out of $n_1^* = 172$).

The same pattern can be seen for Set 2, with slightly better results. Considering the students who submitted both assignments, and who obtained less-than-perfect on the first assignment ($n_2^* = 180$), two-thirds (**0.6556**) did better on the second assignment than they did on the first.

Similarly, for Set 3. Considering the students who submitted both assignments ($n_3 = 187$), two-thirds (**0.6684**) again did better on the second assignment than they did on the first.

Some additional statistics are worth noting.

- Of the students whose grade improved between the first set of assignments ($n_1 = 101$), The majority (**0.7624**) did not do poorer on the second set of assignment.
- Of the students whose grade did not improve between the first set of assignments ($n_2 = 91$), most students (**0.5714**) improved on the second set of assignments.

All this appears to suggest that the students are leaning *something* by completing two versions of the same assignment and having feedback to respond to when completing the second version. Optimistically, we might think that they are improving their skills in writing, reading, analysis, and synthesis. Cynically, we might think that they are improving their skills in understanding how the assignment is graded and what is being asked for—i.e., they are simply doing what the TA asked for. However, I am of the view that being able to understand what is being asked, and responding appropriately, is actually an important skill—both within and without the academy.

Some confounding variables that account for these data can, I think, be ruled out. On the one hand, one might worry about consistency of grading. Perhaps the second version of the assignment was simply graded more leniently. I do not believe that this would explain the data because the grades did not increase *uniformly*—i.e., some students' grades decreased. And, in fact, the average grade for the first and second assignments in the second set were uniformly *lower* than the average grade for the first and second assignments in the first set.

Another worry involves implicit bias. Perhaps there is some unconscious expectation that those who did well on the first component of each assignment set would also do well on the second component. However, this is not a plausible explanation for these data, because all of the assignments were graded anonymously in this course.

A final worry concerns consistency of grading because this course had 7 TAs, so it is possible that one TA graded more lightly than another. However, this possibility is ruled out by two considerations. First, the TAs only graded students that were registered for their tutorial sections, meaning the same students were graded by the same teaching assistant for all four short assignments and two papers. Thus, the only inconsistency might be *between* TA sections. However, for each of the first two sets of assignments, I double-checked all of the TAs' grades to ensure consistency across sections.

4.3 Teaching Evaluations

In addition to soliciting mid-term feedback from my students, I take seriously the feedback that I receive in the year-end evaluations. To ensure these data are as useful as they can be, given the limitations of the medium, I include two bonus marks in my syllabi for a 'course evaluation game'. The following instructions are provided at the start of the semester:

> *Course Evaluations Game***.** If a 3/4 majority of registered students fill out the year-end evaluation, then everyone will receive two bonus marks for the course. Note that this bonus assignment has a structure typical of a *prisoner's dilemma*: If most students cooperate (fill out the evaluation), then it is in your individual interest not to cooperate (because you can get a bonus mark without expending additional effort in filling out the evaluation). Further, if most students defect (fail to fill out the evaluation), it is again in your best interest to defect (otherwise, you would have expended additional effort for nothing). This is a dilemma because it will always be in your own best interest to defect; however, it is in everyone's best interest to cooperate.

This is useful because some research shows that a 2/3 response rate is the minimum standard to inform the ability to present the spread of ratings as adequately meaningful of the class.[1] In Winter 2022, both my courses ($n_1 = 213$, $n_2 = 209$) successfully cooperated on this dilemma, thus obtaining two bonus marks for everyone registered. This means the data obtained in the student evaluations for one semester included more than 300 students.

A summary of the qualitative and quantitative results of my student evaluations are provided below.

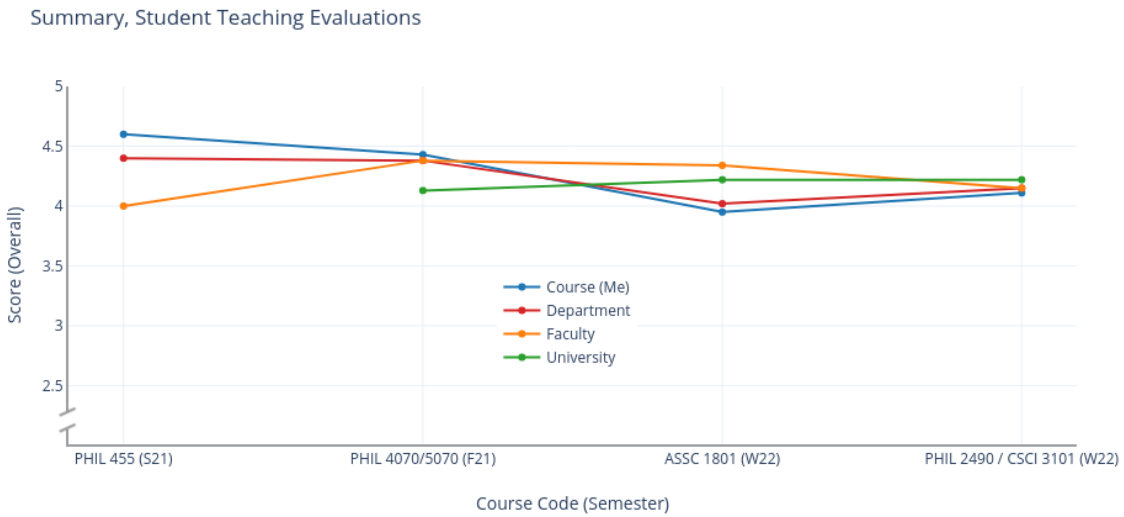
Summary of Teaching Evaluations as Instructor of Record (Quantitative)

The results summarised below are overall results, averaging the scores for the following items under the heading of 'teaching effectiveness':

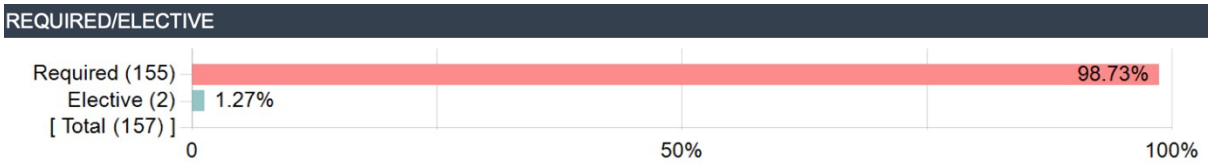
- The instructor conducted the class / clinical in such a way that I was stimulated to learn.
- The instructor organized the class well.
- The instructor communicated clearly during the class.
- The instructor showed enthusiasm for the subject matter of the class.
- The instructor used fair evaluation methods to determine grades.
- The instructor provided constructive feedback (considering the class size).
- The instructor showed genuine concern for my learning.

with responses as follows: 1 – not at all, 2 – somewhat, 3 – moderately, 4 – mostly, 5 – a great deal.

[1] See, e.g., B. G. Davis. 2009. *Tools for Teaching*. John Wiley & Sons.



It is worth noting that, with the exception of ASSC 1801 (Case Studies in Computing and Society), my overall teaching evaluations are consistently on a par with, or in excess of, the averages for each of the department, the faculty, and the university. The course ASSC 1801 is an exception, in part, because the course is restricted to students in the Faculty of Computer Science, but the course is housed in the Faculty of Arts and Social Science, so in this case, comparison with the department and faculty is not meaningful. Specifically, only two (2) students who completed to the course evaluations said that they took this course as an elective:



To give a better sense of responses to individuals questions, the results summarised below are taken from PHIL 455, taught at the University of Toronto (Summer, 2021), which is the most representative of my teaching. (Noting that too few students responded for PHIL 4115/5115 (Fall 2021), and both ASSC 1801 and CSCI 3101 (Winter 2022) are housed outside my home department.)

These results summarised below are in response to the following institutional questions, with responses as follows: 1 – not at all, 2 – somewhat, 3 – moderately, 4 – mostly, 5 – a great deal.

Q1. I found the course intellectually stimulating

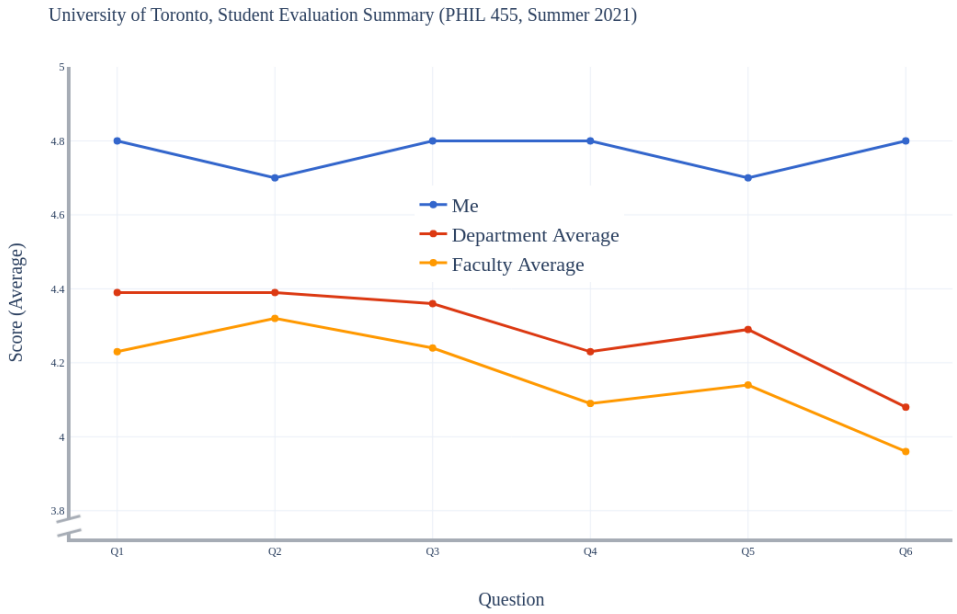
Q2. The course provided me with a deeper understanding of the subject matter.

Q3. The instructor created an atmosphere that was conducive to my learning.

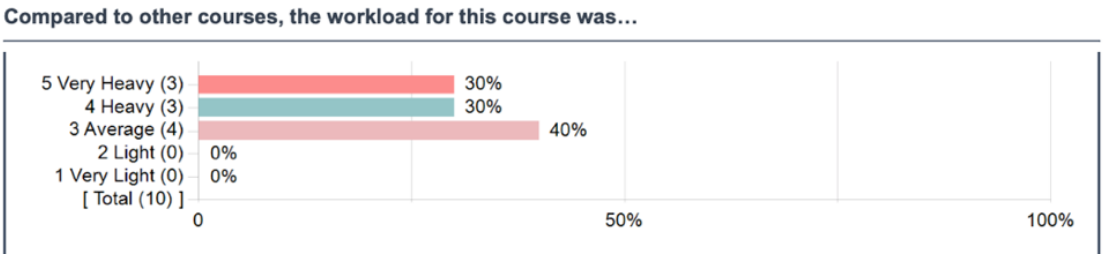
Q4. Course projects, assignments, tests, and/or exams improved my understanding of the course material

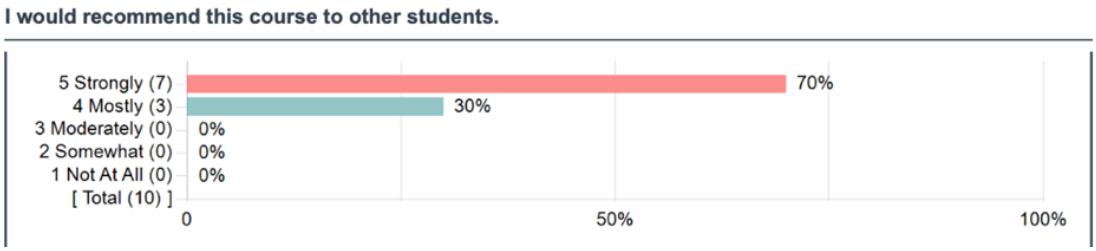
Q5. Course projects, assignments, tests, and/or exams provided opportunity for me to demonstrate an understanding of the course material.

Q6. Overall, the quality of my learning experience in this course was: (1 – poor, 2 – fair, 3 – good, 4 – very good, 5 – excellent).

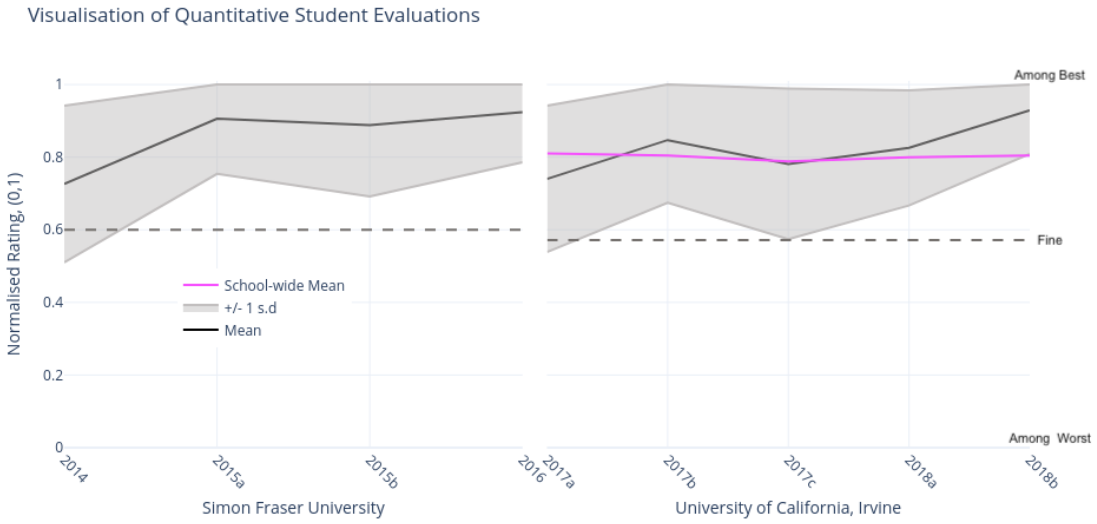


Despite the fact that the majority (0.60) of students registered thought the workload for this course was ‘heavy’ or ‘very heavy’ compared to other courses, the majority of students (0.70) also said they would ‘strongly’ recommend this course to others.





Summary of Teaching Evaluations, as Teaching Assistant (Quantitative)



5- Very Much So 3- So-So 1- Not At All

PHIL xx1: Critical Thinking
Spring, 2016; $n = 34$

	MEAN	St. Dev.
The grading was fair	4.68	0.53
Assignments were returned promptly	4.88	0.32
TA directed the tutorial well	4.56	0.95
Tutorials were helpful in clarifying course material	4.56	0.91
Comments on written work were helpful	4.56	0.77
TA was well prepared	4.71	0.57
TA was easy to follow	4.47	0.81
TA answered questions helpfully	4.79	0.47
TA was aware when students did not understand the material	4.53	0.65
TA had a positive attitude toward thoughtful disagreement	4.74	0.61
Overall (Tutorials)	**4.53**	**0.74**
Overall (TA)	**4.62**	**0.69**

PHIL xx1: Critical Thinking
Fall, 2015; $n = 49$

	MEAN	St. Dev.
The grading was fair	4.37	0.87
Assignments were returned promptly	4.71	0.57
TA directed the tutorial well	4.55	0.76
Tutorials were helpful in clarifying course material	4.39	0.88
Comments on written work were helpful	4.08	1.08
TA was well prepared	4.57	0.70
TA was easy to follow	4.29	0.97
TA answered questions helpfully	4.41	0.99
TA was aware when students did not understand the material	4.16	1.08
TA had a positive attitude toward thoughtful disagreement	4.41	1.05
Overall (Tutorials)	**4.42**	**0.79**
Overall (TA)	**4.44**	**0.98**

PHIL xx1: Critical Thinking
Spring, 2015; $n = 37$

	MEAN	St. Dev.
The grading was fair	4.00	1.01
Assignments were returned promptly	4.69	0.57
TA directed the tutorial well	4.27	0.98
Tutorials were helpful in clarifying course material	4.41	0.85
Comments on written work were helpful	3.73	1.06
TA was well prepared	4.35	0.91
TA was easy to follow	4.24	1.00
TA answered questions helpfully	4.30	0.90
TA was aware when students did not understand the material	4.32	0.99
TA had a positive attitude toward thoughtful disagreement	4.41	0.94
Overall (Tutorials)	**4.35**	**0.81**
Overall (TA)	**4.53**	**0.76**

PHIL 120W: Introduction to Ethics
Fall, 2014; $n = 30$

	MEAN	St. Dev.
The grading was fair	3.17	1.13
Assignments were returned promptly	4.30	0.74
TA directed the tutorial well	3.60	0.95
Tutorials were helpful in clarifying course material	3.50	1.02
Comments on written work were helpful	3.73	0.96
TA was well prepared	4.03	0.84
TA was easy to follow	3.33	1.14
TA answered questions helpfully	3.57	1.12
TA was aware when students did not understand the material	3.20	1.01
TA had a positive attitude toward thoughtful disagreement	3.48	1.28
Overall (Tutorials)	**3.63**	**0.91**
Overall (TA)	**3.63**	**1.08**

7- Among Best 4- OK 1- Among Worst

LPS 31: Introduction to Inductive Logic
Spring, 2018; $n = 22$

	MEAN	St. Dev.	MEAN (All S.S. Courses)
TA was competent in course material	6.55	0.66	5.91
TA was able to make presentations clearly	6.48	0.73	5.85
TA was responsive to students	6.45	0.72	5.97
TA was able to integrate the lecture and discussion material	6.50	0.72	5.94
TA was present and on time for discussion sections and office hours	6.64	0.57	6.09
The discussion sections were useful to the success of the course	6.36	1.02	5.81
I would expect another course with this TA to be	6.64	0.57	5.84
General teaching effectiveness	**6.50**	**0.84**	**5.63**

LING 3: Introduction to Linguistics
Winter, 2018; $n = 75$

	MEAN	St. Dev.	MEAN (All S.S. Courses)
TA was competent in course material	5.84	1.14	5.83
TA was able to make presentations clearly	5.87	1.10	5.78
TA was responsive to students	5.93	1.14	5.94
TA was able to integrate the lecture and discussion material	5.95	1.15	5.89
TA was present and on time for discussion sections and office hours	6.07	1.11	6.11
The discussion sections were useful to the success of the course	5.74	1.36	5.74
I would expect another course with this TA to be	5.85	1.24	5.78
General teaching effectiveness	**5.78**	**1.11**	**5.60**

LING 51 / PSYC 56: Acquisition of Language
Fall, 2017; $n = 35$

	MEAN	St. Dev.	MEAN (All S.S. Courses)
TA was competent in course material	5.60	1.57	5.72
TA was able to make presentations clearly	5.83	1.52	5.65
TA was responsive to students	5.66	1.57	5.83
TA was able to integrate the lecture and discussion material	5.45	1.63	5.77
TA was present and on time for discussion sections and office hours	5.50	1.61	5.98
I would expect another course with this TA to be	5.60	1.55	5.65
General teaching effectiveness	**5.47**	**1.45**	**5.52**

LPS 31: Introduction to Inductive Logic
Spring, 2017; $n = 26$

	MEAN	St. Dev.	MEAN (All S.S. Courses)
TA was competent in course material	6.19	1.44	5.86
TA was able to make presentations clearly	5.88	1.55	5.79
TA was responsive to students	6.16	1.46	5.93
TA was able to integrate the lecture and discussion material	6.08	1.47	5.89
TA was present and on time for discussion sections and office hours	6.26	1.45	6.06
The discussion sections were useful to the success of the course	5.87	1.75	5.74
I would expect another course with this TA to be	6.00	1.56	5.77
General teaching effectiveness	**5.93**	**1.21**	**5.63**

ECON 15A: Probability & Statistics for Economics I
Winter, 2017; $n = 49$

	MEAN	St. Dev.	MEAN (All S.S. Courses)
TA was competent in course material	5.02	1.58	5.86
TA was able to make presentations clearly	5.02	1.68	5.78
TA was responsive to students	5.31	1.75	5.96
TA was able to integrate the lecture and discussion material	5.06	1.69	5.90
TA was present and on time for discussion sections and office hours	5.73	1.55	6.15
The discussion sections were useful to the success of the course	4.90	1.92	5.72
I would expect another course with this TA to be	4.88	1.75	5.79
General teaching effectiveness	**5.18**	**1.411**	**5.67**