2013

Response Rates: A case study based on observed variability in UBC student ratings of instructors

ABDEL AZIM ZUMRAWI & MARIANNE SCHROEDER

Prepared for the SEoT Committee, July 2013

Summary

In Student Evaluation of Teaching (SEoT), it is important to ensure, within a reasonable degree of confidence, that the responses are a sufficient representation of the students in the class. This is necessary in order for the results to be generalized and interpreted in a meaningful way. This report addresses the factors that influence and are used in the determination of required or desirable response rates. For a given level of confidence and margin of error, the required response rate depends on the underlying variability in the student ratings of instructors. At UBC, and confirmed in the literature, students tend to select higher ratings more often than lower ratings; however inter-institutional differences could exist. In this report, simulations are provided to compare what the "statistically adequate response rate" is under a wide range of scenarios.

The more the students agree in their ratings, the lower the required response rate. The variability in the UBC student ratings of instructors for the period 2009 – 2012 was examined. Overall, 77% of UBC students give favourable instructor ratings (4 or 5 on a scale of 1-5). This estimate is higher than the 70% used to calculate the recommended response rates reported by Hakstian in the 2010 report. A theoretical consequence of this difference in variability is that the recommended required response rates, for a given class size, confidence level and margin of error, will be slightly higher than actually required. However, this difference needs to be quantified.

Based on the observed UBC variability, new estimates of the minimum recommended response rates were calculated for a range of class sizes, confidence levels and margins of error. These recommended response rates are grouped into ten class size categories. If the response rate for a given class size is below the recommended rate, the evaluation data should be interpreted with care because it may not be generalizable.

Finally, the 2012 UBC response rates were summarized and compared to the minimum recommended rates. The majority of sections with 50 or more students, representing about 65% of total enrolment, met or exceeded the minimum recommended rates. However, less

than half of the sections with 50 or less students, representing 35% of total enrolment, met or exceeded the minimum recommended rates.

Introduction

An important aspect of SEoT is to ensure, within a reasonable degree of confidence, that the respondents represent the students in the class. This is necessary in order for the results to be generalized. Hakstian, in his 2010 report to the SEoT committee, compared UBC's overall response rates to "McGill's Acceptable Response Rates" and to response rates recommended by Nulty (2008). The latter are based on two scenarios computed under what Nulty termed "liberal conditions" and "stringent conditions". The "liberal conditions" included a 10% sampling error and an 80% confidence level, i.e., results from surveys using these response rates will be accurate within +/- 10%, 8 times out of 10. The stringent conditions included a 3% sampling error and a 95% confidence level, i.e., results will be accurate within +/- 3%, 19 times out of 20. Both of Nulty's scenarios used an estimate of 70% of favourable instructor ratings (4 and 5 on a 1-5 scale), and are obtained from data covering five years in one (unnamed) Australian University. Before we compare UBC response rates to Nulty's (2008) recommended rates, we need to determine if the variability in the UBC student ratings is different than the Australian university i.e., how much do UBC students agree on their instructor rating compared to the 70% reported from the Australian university?

The objectives of this case study are:

1) to discuss the factors used in determining desired or minimally acceptable response rates in student evaluations of teaching;
2) to estimate the variability in the UBC student ratings of instructors based on data covering a 4-year period (2009 – 2012); and,
3) to calculate estimates of the minimum required response rates based the observed UBC variability.

Factors for the Determination of Required Response Rates

The required response rate depends on the number of students in the class, the variability in student ratings, and the margin of error and confidence level desired. These factors are briefly discussed below.

Underlying variability in students ratings of instructors

In the hypothetical, but unlikely, case where all the students in a class share the same opinion of their instructor (whether it be favourable or unfavourable), it could be argued that a single response from one student would suffice as a representation of the class. However, as students typically "vary" in their evaluation of the instructor, more responses are needed to more accurately represent the class. The first factor in determining desired response rates is an estimate of the underlining variability in the student responses. The highest level of "variability in responses" occurs when students are split 50:50 in their rating of their instructor. This is the most conservative scenario. However, in teaching evaluations, students tend to select higher ratings (4 and 5 on a scale of 1-5) more frequently than lower ratings (1, 2 and 3).

Margin of Error

The margin of error determines the degree of accuracy with which to estimate student ratings of instructors. For example, if 75% of the responding students in a class give their instructor a favourable rating, how accurate is this estimate in representing the whole class? Do we want the results to be within +/- 3%, 5% or 10% and what are the implications of this choice? The choice of margin of error, coupled with the choice of a confidence level (to be discussed below), will affect how we interpret the results of the student evaluation of teaching. This in turn impacts our interpretation of the estimated mean of the six University Module Items (UMIs).

Confidence Level

An 80% confidence level would mean that the estimated rating is accurate within a given margin of error 8 times out of 10. A 95% confidence would mean that the results are accurate 19 times out of 20. In the above example where 75% of the students in a class give their instructor favourable ratings; and given a 5% margin of error and a 90% confidence level, the students' approval is estimated to be between 70% and 80% (75% ± 5%), 9 times out of 10.

A Simulation of the Effect of Underlying Variability on Required Response Rates

To demonstrate the effect of underlying student variability on desired response rates, the desired response rates were calculated for a range of class sizes, margins of error (3%, 5% and 10%), and confidence levels (80%, 90% and 95%) for four scenarios of underlying variability in ratings (70:30; 75:25; 80:20 and 90:10 splits). A 70:30 split would mean that 70% of the students gave favourable ratings to their instructor, while 30% gave an unfavourable rating.

For example, in this simulation, using a 5% margin of error and a 90% confidence level, the required responses for a class of 200 students were found to be 106, 100, 92 and 66, for the four variability scenarios, respectively. This shows that the more students agree on their instructor ratings, the lower would be the required response rate. In this case, the 90:10 split resulted in the least required response rate to meet the specified 5% margin at a 90% confidence.

To illustrate the effect of underlying variability on response rates, the desired rates for the 70:30 and 90:10 scenarios are shown in figures 1 and 2, respectively.

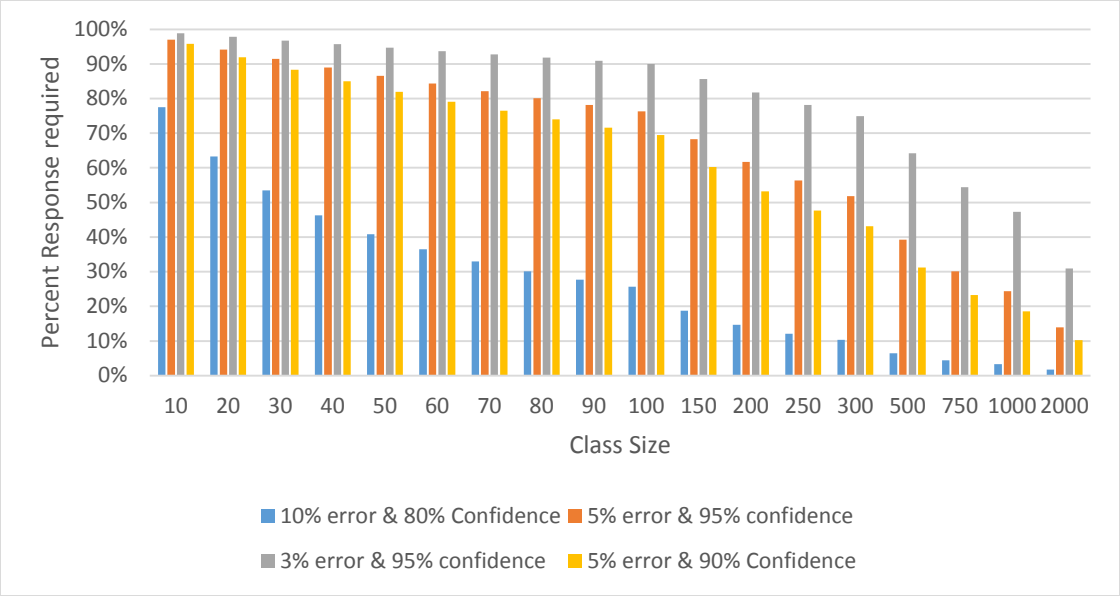


Figure 1: Required response rates for an underlying 70:30 variability in students rating

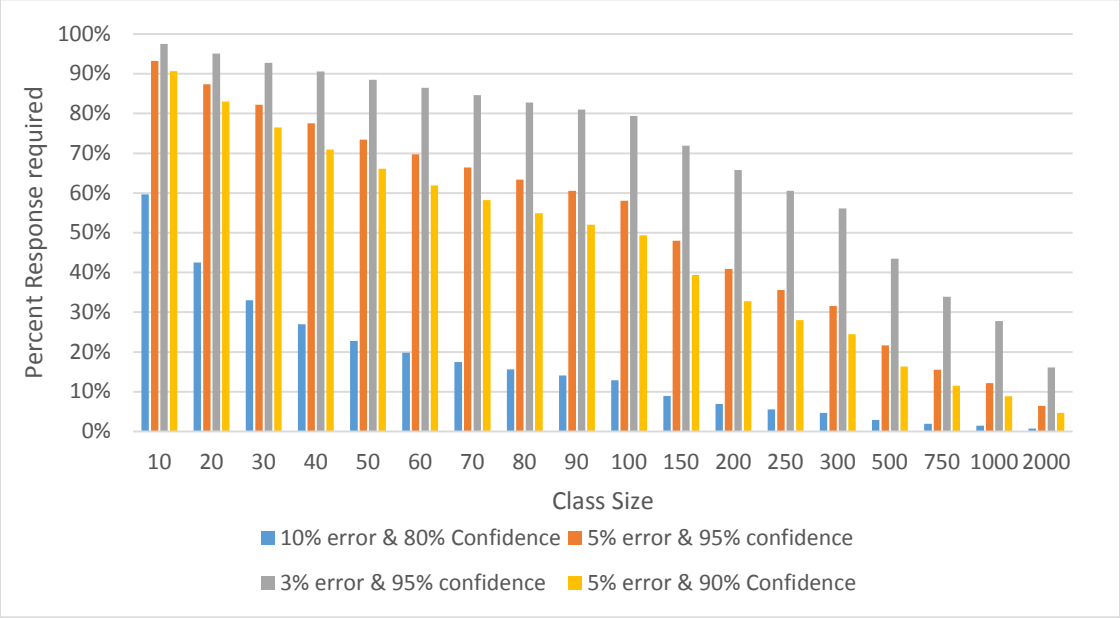


Figure 2: Required response rates for an underlying 90:10 variability in students rating

This simulation exercise shows that small differences in the underlying variability of student ratings may not affect the desired response rates significantly. However, if variability in the UBC students rating is significantly different from the 70:30 split used in Nulty's study, and recommended by Hakstain (2010), new estimates of what is statistically acceptable should be calculated based on the observed UBC variability.

Variability in the UBC student Ratings of Instructors (2009-2012)

The UBC University Module Items (UMI) are six questions mandated by the university's Student Evaluation and Teaching policy. The six UMI are intended to solicit student responses concerning the instructor's: clarity of expected learning outcomes, communication of subject matter; inspiring interest in the subject, fairness in evaluation, concern for students learning, and overall instructor effectiveness. For example, the sixth UMI states "Overall, the instructor was an effective teacher". For each UMI statement, students are presented with five multiple choices ranging from "Strongly Agree" to "Strongly Disagree" with a "neutral" option, making it an evaluation on a scale of 1 to 5.

Under a binomial distribution model, each student's response could be classified as being favourable or unfavourable to the instructor, without much loss of information. In this case, the "strongly agree" and "agree" responses are favourable, the "strongly disagree", "disagree" and "neutral" are unfavourable ratings.

The average percent favourable responses for the six UMI from 2009 through 2012 are given in Table 1. The averages for UMI question 6 (UMI_6) were 75%, 77%, 77% and 77%, for the four years, respectively. The variability in the six UMIs has been consistent over the four years with an overall average favourable response of 77%. These results are consistent with findings by Hakstian (2010) on the stability of the mean UMI scores.

Table 1: Variability in student ratings of instructors for the 6 University Module Items (UMI) for years 2009 -2012.

Year / University Module Item	Number of Responses	% Favourable Responses (4-5)
2009		
UMI_1	156,885	79%
UMI_2	156,702	74%
UMI_3	156,467	69%
UMI_4	144,360	72%
UMI_5	156,652	77%
UMI_6	156,577	75%
2010		
UMI_1	167,747	80%
UMI_2	167,452	76%
UMI_3	167,121	72%
UMI_4	155,604	74%
UMI_5	167,408	80%
UMI_6	167,356	77%
2011		
UMI_1	169,404	80%
UMI_2	169,167	76%
UMI_3	168,832	72%
UMI_4	156,972	75%
UMI_5	169,126	80%
UMI_6	168,790	77%
2012		
UMI_1	167,426	80%
UMI_2	167,093	76%
UMI_3	166,785	72%
UMI_4	156,158	75%
UMI_5	167,150	80%
UMI_6	166,848	77%

The variability estimates in Table 1 show that the 70:30 split (from an Australian university) used to calculate Nulty's recommended response rates (Hakstian, 2010), overestimate the variability in the UBC students rating. This overestimation of variability leads to inflated required response rates for a given class size, confidence level and margin of error (Figure 1).

For example, in a class of 100 students, the required responses based on the Australian estimate of variability and that of UBC are 52 and 46, respectively.

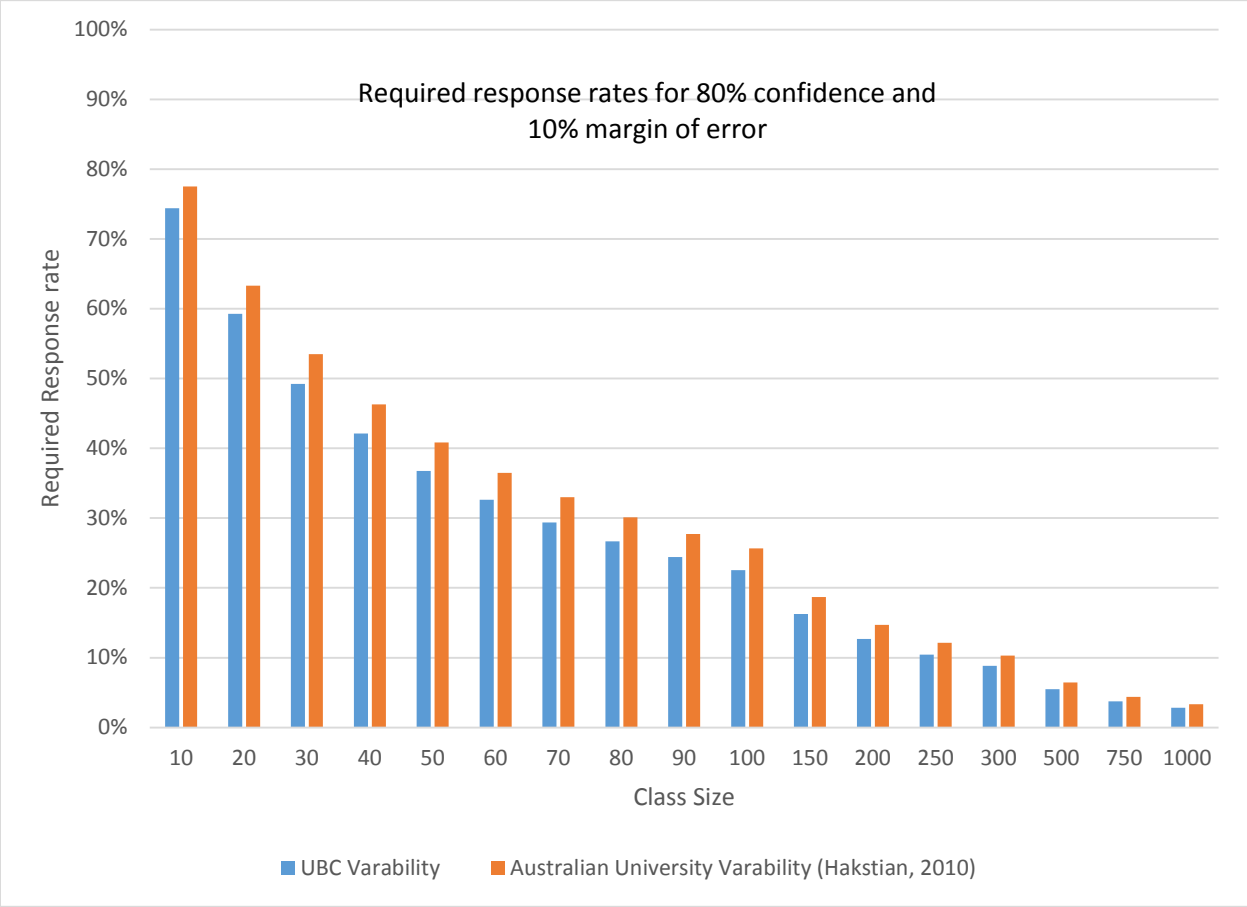


Figure 3: Comparison of required response rates based on variability in student ratings from an Australian university (Hakstian, 2010) and those based on UBC data.

Required Response Rates Based on Observed UBC Variability in Student Ratings

Table 2 shows the required response rates, based on the observed variability in the UBC student ratings. These rates were calculated for combinations of class sizes (10 – 1000), margins of error (± 5% or ± 10%) and confidence levels (80% or 90%) using sample size computation techniques (Dillman, 2009), based a binomial distribution.

Table 2: Required response rates for combinations of confidence level and margin of error based on variability in the UBC students rating of instructors.

	Required response rates based on variability in UBC ratings	
Class Size	**80% Confidence & ± 10%**	**90% Confidence & ± 5%**
10	74%	95%
20	59%	91%
30	49%	86%
40	42%	83%
50	37%	79%
60	33%	76%
70	29%	73%
80	27%	71%
90	24%	68%
100	23%	66%
150	16%	56%
200	13%	49%
250	10%	43%
300	9%	39%
500	5%	28%
750	4%	20%
1000	3%	16%

.

The response rates in Table 2 were summarized in Table 3, by categories of class sizes. The class size categories are based on the 2012W class size distribution. Ratings from response rates below the recommended minimum should be interpreted with care as they cannot be generalized with an acceptable degree of confidence or within a reasonable margin of error, particularly if these evaluations differ from previous years.

Table 3: Recommended minimum response rates

	Recommended Minimum Response Rates based on 80% confidence & ± 10% margin
Class Size	
< 10	75%
11 - 19	65%
20 - 34	55%
35 - 49	40%
50 - 74	35%
75 - 99	25%
100 - 149	20%
150 - 299	15%
300 - 499	10%
> 500	5%

Analysis of the 2012 Response Rates

The 2012W response rates for all sections, across all Faculties were summarized by class size categories and compared to the minimum recommended rates from Table 3. The results are presented in Table 4.

In 2012W, the majority of sections with 50 or more students, representing 65% of total enrolment, met or exceeded the minimum recommended response rate. Of these, all sections with 150 or more students met or exceeded the minimum recommended response rates. In fact, the response rates in many of the big sections would exceed the required rates even at more stringent conditions in terms of confidence level and margin of error. However, only 50% or less of sections with under 20 students met or exceeded the minimum required rates. While this category represents 35% of sections evaluated, it contains 9% of total enrolment in 2012W.

The distribution of the 2012W response rates (based on data in the online system) by Faculty is given in Appendix 1.

Table 4: Summary of the 2012W response rates shown the percentage of sections that met or exceeded the recommended rates[1]

Class Size	Sections evaluated	Enrolment	% sections meeting or exceeding the recommended response rate
≤ 10*	718	5,438	38%
11 -19	1206	18,265	30%
20 -34	1448	37,937	50%
35 - 49	725	29,204	60%
50 -74	521	31,325	85%
75 -99	320	27,740	95%
100 -149	322	39,913	95%
150 - 299	330	69,246	100%
300 - 499	30	10,376	100%
> 500	1	607	100%

[1] course sections with < 5 students were excluded from this analysis.

Conclusions

The recommended response rates presented in this report (Table 3) are based on the estimated variability in the UBC student ratings of instructors for the years 2009-2012

In 2012W, the majority of sections with 50 or more students, representing 65% of total enrolment, met or exceeded the minimum recommended rates. Efforts to increase students' participation in online surveys could thus be focused on classes with under 50 students.

The recommended rates could be incorporated into future high level summaries of student evaluations of teaching to highlight course sections that have not met the minimum recommended rates.

References

Dillman, D.A. 2009. *Mail and internet surveys: the tailored design method.* Wiley.

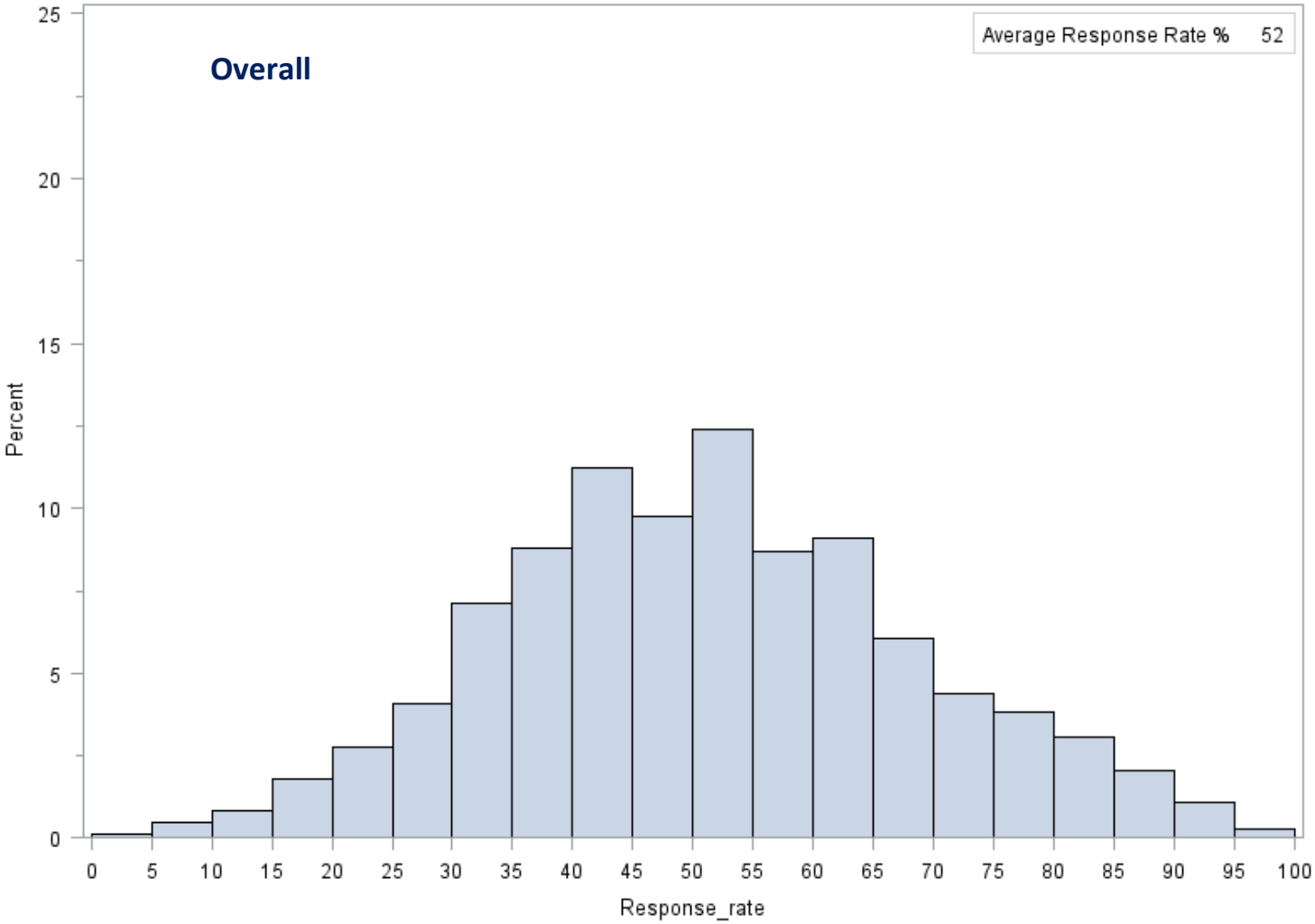
Hakstian (2010). Student Evaluation of Teaching: Response rates. *Report for SEoT Committee, University of British Columbia*.

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment and Evaluation in Higher Education, 33, 301-314*

Appendix 1

Distribution of the 2012W response rates

Overall


Average Response Rate % 52
Percent
25
20
15
10
5
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Response_rate