

Timing of Student Evaluations of Teaching in a Veterinary Medical Education Setting

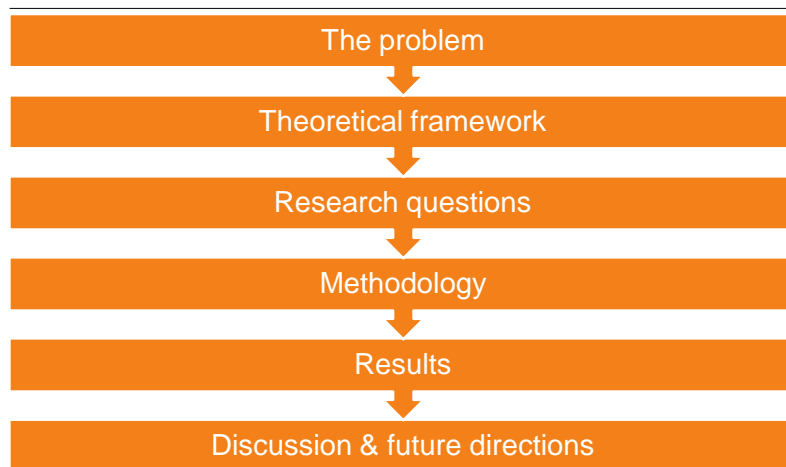
Misty Bailey,^{1,2} MA, ELS(D); Dianne Mawby,¹ DVM, MVSc, DACVIM

¹College of Veterinary Medicine, University of Tennessee

²Department of Educational Leadership & Policy Studies, University of Tennessee



Outline

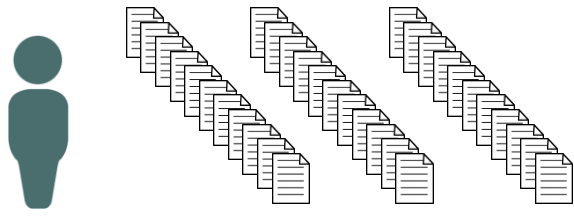


THE PROBLEM

Student Evaluations of Instruction

> 30 mean evaluations per student per semester

Why 30?



What do students evaluate?

- Semester curriculum
- 9 – 11 Courses
 - Specialty team taught
- Instructors
 - 4 lectures/labs taught
 - Minimum for evaluation



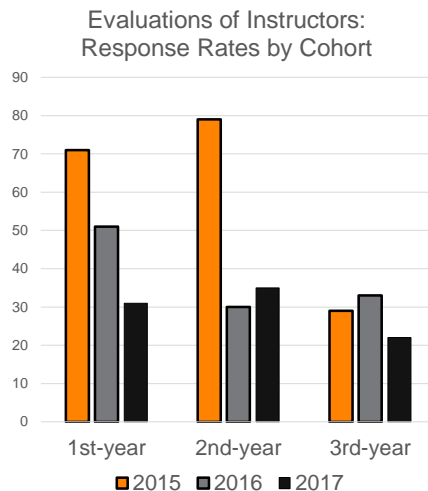
Ex: VMD 846–Multispecies Medicine



Low Evaluation Response Rates

- Survey fatigue¹
 - 11 or more = lower response rate
- Validity of results

1. Adams & Umbach, 2010



Timing of Evaluations

- **After final exams?**
 - Knowledge of grade^{1,2}
 - Negative bias³
 - Fewer substantial comments⁴

1. Arnold, 2009
2. Hoefler, Yurkiewicz, & Byrne, 2012
3. Cho & Cho, 2017
4. McNulty, et al., 2010



THEORETICAL FRAMEWORK

Study Theoretical Framework

- Self-serving bias¹
 - Internalize success
 - Externalize failure
- Revenge theory^{1,2}
 - Enact revenge due to poor performance³
 - Provide poor evaluation scores, negative, non-constructive comments



1. Arnold, 2009
2. Cho & Cho, 2017
3. Maurer, 2006

RESEARCH QUESTIONS

Research Questions

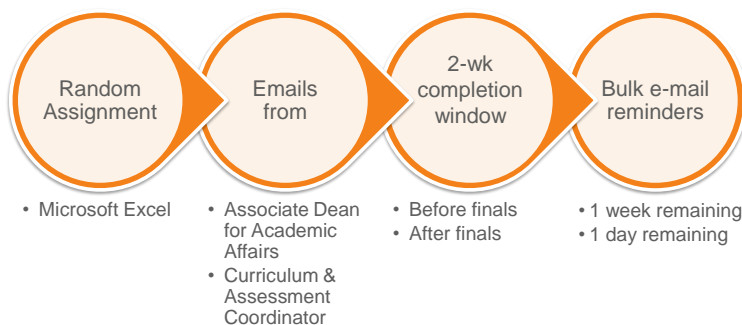
Does the timing of veterinary students' evaluations (before or after final exams) of instructors affect the following, as related to self-serving bias and revenge theories:

1. Completion rate?
2. Numerical scores on evaluations?
3. Number and substance of comments?

METHODOLOGY

Student Evaluation Methods

Fall 2017 & Spring 2018



Research Site & Population

- UT College of Veterinary Medicine
 - All veterinary students in first 3 years of study
 - 264 students
 - 2,483 evaluations
 - 4th-year students excluded

- IRB Approved



Quantitative Data CLEANING¹–Evaluation Items

- Non-normal distribution
 - Robust with large sample size, Likert scale, *t* test, conservative alpha²

- Missing data
 - Most items <5%
 - Large sample size
 - Pairwise deletion



1. Morrow & Skolits, 2017
2. Garson, 2012



Quantitative Data CLEANING

- **Quantitative comment data**
 - Non-normal distribution
 - Kurtosis = 3.10
 - Robust with independent t test¹



1. Garson, 2012

Quantitative Data ANALYSIS (SPSS 24)

- | | |
|--|--|
| <ul style="list-style-type: none"> • Completion Rate <ul style="list-style-type: none"> – 2016/17 & 2017/18 – Chi-square • Evaluation Items <ul style="list-style-type: none"> – Independent t tests | <ul style="list-style-type: none"> • Comments <ul style="list-style-type: none"> – Word Count <ul style="list-style-type: none"> • Independent t test – Comments Provided/
Evaluations
Completed <ul style="list-style-type: none"> • Chi-square |
|--|--|

Qualitative Data ANALYSIS¹ (NVivo 12)

- Team eclectic **coding**
 - Descriptive (exploratory)
 - Magnitude
- Developed **categories** from 1st-round descriptive coding
- Identified **themes**

1. Saldaña, 2013



UTCVM
INSTITUTE OF AGRICULTURE
UNIVERSITY OF CALIFORNIA

RESULTS

1. Completion rate

- Instructors only

- $\chi^2(1) = 129.95$

- $p < .001$

- **Significantly lower completion rate in 2017/18**

Semester	Completion Rate
Fall 2016	47%
Spring 2017	23%
Fall 2017	36%
Spring 2018	23%



2. Item numerical ratings

- Instructors only

- **No statistical differences in any item before vs after finals**

Item #	Increased means <u>before</u>	Item #	Increased means <u>after</u>
2	Lectures/labs organized, clear objectives (+.05)	1	Instructor concerned with facilitating learning (+.01)
3	Instructor willing to discuss material outside class (+.01)	8	Instructor covered material at pace reasonable for me (+.06)
4	Instructor encouraged interaction, answered questions in class (+.03)	9	Directions for assignments clear, specific (+.01)



3. Percent of Evaluations with Comments*

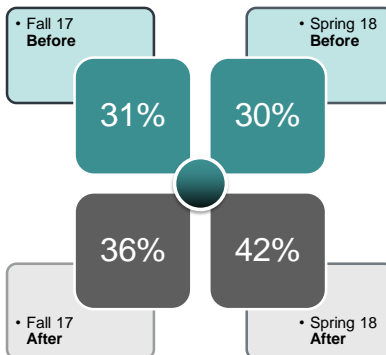
- $\chi^2(1) = 10.00$

- $p = .002$

- Overall

- Before: 30%

- After: 40%



3. Mean word count per comment*

- $t(258) = 2.33$

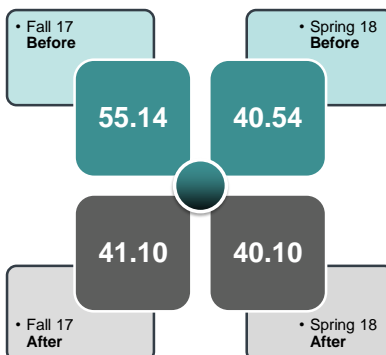
- $d = .26$

- $p = .02$

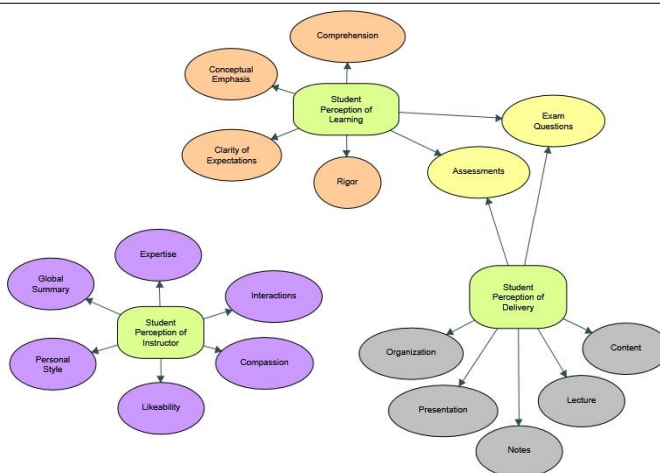
- Overall:

- Before: 49.59 ± 39.06

- After: 40.47 ± 30.59



3. Student comments – Categories*



3. Student Comments – PRESENTATION Themes*

- Before Finals
 - “Entertaining” instructor
 - Satisfactory delivery pace
 - Slides lack detail
- After Finals
 - Interesting & enthusiastic instructor
 - Slow delivery pace
 - Slides & notes lack coordination
 - Slides & notes overwhelming

3. Student Comments – ASSESSMENT Themes*

- Before Finals
 - Material presented in too much detail to study
- After Finals
 - Information tested incongruent with material given in class
 - Low-stakes assessments appreciated
 - Exams challenging but fair

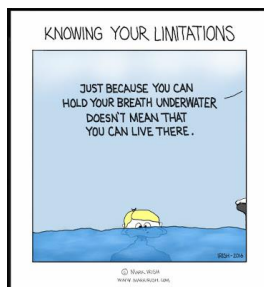


*Year 3 cohort ONLY

DISCUSSION & FUTURE DIRECTIONS

Limitations

- Same (anonymous) participants filled out multiple evaluations for certain instructors who taught in more than one course
- Restricted to one veterinary school



UTCVM
INSTITUTE OF AGRICULTURE
THE UNIVERSITY OF TORONTO

Practical Uses

- Can we use results to:
 - Inform change
 - Alter current practices
 - Increase the value of student evaluations to instructors
 - Reduce student stress



UTCVM
INSTITUTE OF AGRICULTURE
THE UNIVERSITY OF TORONTO

Future Directions

- Analysis of 1st & 2nd-yr student comments
- Examine other higher education student populations



UTCVM
INSTITUTE OF AGRICULTURE
THE UNIVERSITY OF TEXAS AT COLUMBIA

Questions?

Advice?

Acknowledgements:
Dr. Sonya McNeely
Dr. India Lane



Gigi, 13 years old

UTCVM
INSTITUTE OF AGRICULTURE
THE UNIVERSITY OF TEXAS AT COLUMBIA

References

- Adams, M. J. D., & Umbach, P. D. (2010). Nonresponse and online student evaluations of teaching: Understanding the influence of salience, fatigue, and academic environments. *Research in Higher Education, 53*, 576–591.
- Arnold, I. J. M. (2009). Do examinations influence student evaluations? *International Journal of Educational Research, 48*, 215–224.
- Cho, D., & Cho, J. (2017). Does more accurate knowledge of course grade impact teaching evaluation? *Education Finance and Policy, 12*, 224–240.
- Garson, D. (2012). *Testing statistical assumptions*. Raleigh, NC: Statistical Associates Publishing.
- Hoefer, P., Yurkiewicz, J., & Byrne, J. C. (2012). The association between students' evaluation of teaching and grades. *Decision Sciences Journal of Innovative Education, 10*(3), 447–459.
- Maurer, T. W. (2006). Cognitive dissonance or revenge? Student grades and course evaluations. *Teaching of Psychology, 33*, 176–179.
- McNulty, J. A., Gregory, G., Chandrasekhar, A., Espiritu, B., Hoyt, A., Ensminger, D. (2010). Are online student evaluations of faculty influenced by the timing of evaluations? *Advances in Physiology Education, 34*, 213–216.
- Morrow, J. A., & Skolits, G. (2017). *Twelve steps of quantitative data cleaning: Strategies for dealing with dirty evaluation data*. Workshop presented at the American Evaluation Association, Washington, D.C.
- Saldaña, J. (2013). *The coding manual for qualitative researchers*. (2nd ed.). Thousand Oaks, CA: Sage.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Upper Saddle River, NJ: Pearson.