Understanding Course Characteristics and Response Rates: A Path to Improving Courses and Use of Course Evaluation Results

Seungwon Chung, Stephanie Klein, & Aurore Phenow
University of Minnesota

August 4, 2021
Presentation at BlueNotes Global Virtual Experience Conference

UNIVERSITY OF MINNESOTA
Driven to Discover®

INTRODUCTION

Student Evaluations of Teaching at UMN

Centralized administration

- Policy (Vice Provost for Faculty and Academic Affairs)
  - Core content
  - Data collection windows
  - High-level guidance on use of data
- Administration (Office of Measurement Services)
  - Administer program according to policy/guidelines
  - Facilitate exception requests when appropriate

Decentralized

- Actual use of results for improving instruction, tenure decisions

Prior UMN Research

Very little:

Limited resources within OMS + policy owner reluctance to release data to other requestors without a larger research plan

Until:

Grant program facilitating faculty research partnership + more consensus among stakeholders re: research questions

Purpose: Foundational Research

- Course evaluation results are complex, and other factors may be masked
- Beliefs and assumptions can reduce trust in evaluation results (e.g., “nobody likes morning classes”)
- The more we know about student, course, and instructor characteristics in relation to how students evaluate courses, the better future research can be

METHODS

Data

Fall 2019

- Pre-COVID
- Classes were comprised of predominately in-person instruction
- Paper and online SET forms administered
- Final case count (with missingness): 45,379 (1,747 courses & 43,579 students)

Fall 2020

- During COVID
- Classes were comprised of predominately online instruction
- Only online SET forms were administered
- Final case count (complete data only): 35,745 (2,571 courses & 11,939 students)

Measures: Outcomes

- **Student Evaluations of Teaching (SET)**
  - SET have two sections: instructor and course
  - Consists of 6 instructor items and 7 course items
  - 6-point Likert-type scale (*Strongly Agree, Agree, Somewhat Agree, Somewhat Disagree, Disagree, Strongly Disagree)*
  - Fall 2019 and Fall 2020 forms contain the same items except:
    - Fall 2019 item “The course site was easy to use” was dropped
    - Fall 2020 addition: “The amount of effort needed to be successful in this course is reasonable”

Measures: Predictors

Course Predictors

- Gateway/non-gateway
- STEM/non-STEM
- Class size (total enrollment)
- Drop/fail/withdrawal (DFW) rate
- Class time (morning, afternoon, evening)
- Course level (graduate/undergraduate)
- Instruction mode (completely online, primarily online, partially online, face-to-face)

Response Rate

- Percentages based on the number of students who responded out of those enrolled in the course at the time of evaluation

Measures: Covariates

Student Characteristics

- International/domestic
- First generation
- Minority
- Gender (female/male)
- PELL recipient
- Composite ACT score
- High school GPA
- Current GPA
- Cumulative GPA
- Pass/no pass grade in course
- Credits taken

Instructor Characteristics

- Tenured/non-tenured
- Years of teaching at the university
- Doctorate/non-doctorate
- Gender (female/male)
- Minority
- International/domestic

Analysis

1. Data Exploration
   - Descriptive statistics for all variables
   - Descriptive statistics for SET items

2. Fitting the item response theory (IRT) model
   - Basic IRT model (unidimensional vs. multidimensional)
   - Cross-classified multilevel IRT model
   - Fit the models separately for 2019 and 2020 data

Analysis

- **A (multidimensional) cross-classified multilevel item response theory (IRT) model**
  - Why IRT
    - Item-level data
  - Why cross-classified multilevel IRT
    - Item responses are nested within students; Item responses are nested within course; students and courses are not nested but crossed
  - Why multidimensional IRT
    - Students respond to questions about the **course** and the **instructor**

Analysis

- **Multilevel cross-classified rating scale model** (Huang, Chung, & Cai, in preparation)
  - Rating scale model (RSM) (Andrich, 1978a, 1978b)
    - Appropriate for modeling Likert response scale data
  - Extension to cross-classified data structure
    - Adopting a novel parameterization of the nominal response model (NRM; Bock, 1972; Thissen, Cai & Bock, 2011; Thissen& Cai, 2016).
  - Estimation: Metropolis-Hastings Robins-Monroe (MH-RM) algorithm using flexMIRT (Cai, 2015)

Analysis

- The probability that student j (j = 0, … , J) responds k (k = 0, … , K) to item i (i = 0, … , I) in his or her evaluation to course g (g = 0, … , G) is

$$P\left(y_{ijg} = k \middle| \eta_j, \xi_g; a^c, a^r, \boldsymbol{a}^s, \boldsymbol{c}_i\right) = \frac{\exp(z_{ijgk})}{\sum_{m=0}^{K-1} \exp(z_{ijgm})}$$

where $z_{ijgk} = a^c a_k^s \eta_j + a^r a_k^s \xi_g + c_{ik}$

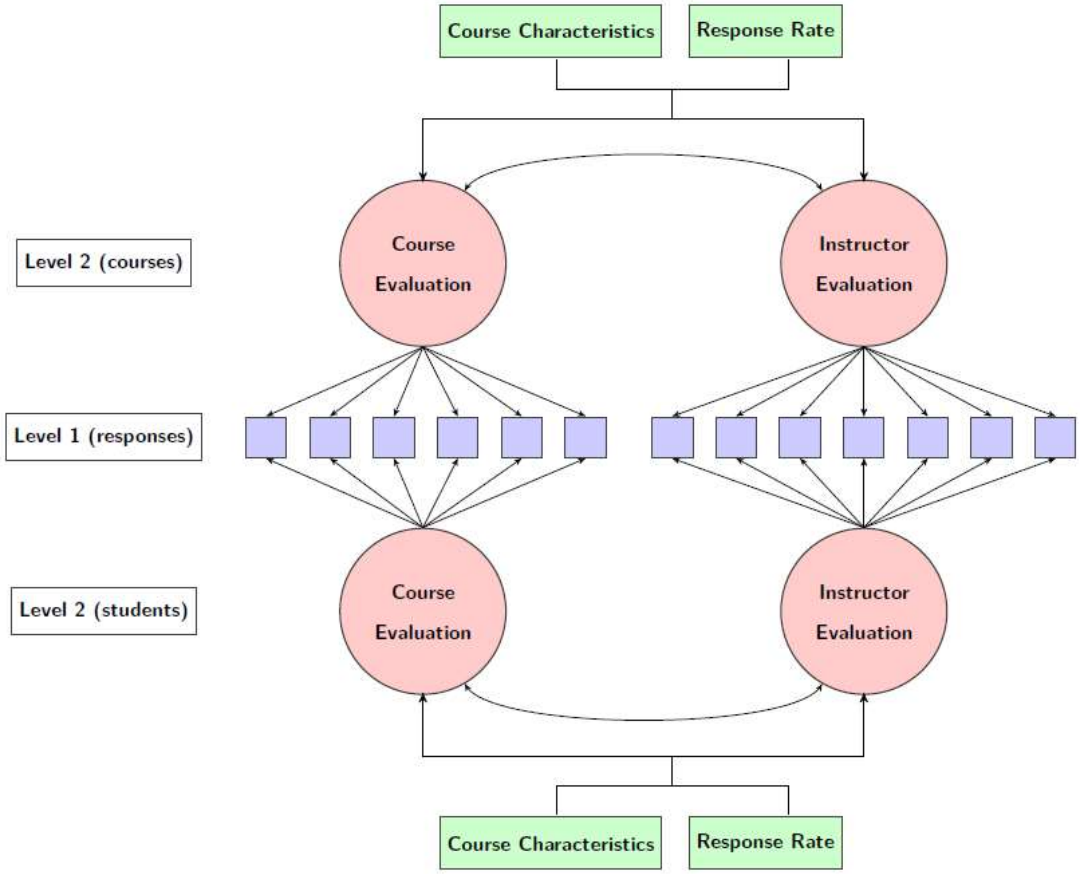
- η_j: student latent variable
- ξ_g: course latent variable
- a^c, a^r: item slopes, $\boldsymbol{a}^s = (0,1,2,3,4)$: scoring function vector
- $\boldsymbol{c}_i$: item i's slope

Initial Analysis

- Model 1: a null model or unconditional model, i.e., a cross-classified multilevel RSM without predictors
- Model 2: a conditional model with course characteristics
- Model 3: a conditional model with response rate
- Model 4: a conditional model with course characteristics and response rate
- Model 5: Model 4 + interaction terms b/w course characteristics and response rate
- Model 6: Model 5 + student and instructor covariates

Initial Analysis

Figure 1. A multidimensional cross-classified multilevel IRT model examining the effects of course characteristics and response rates on student evaluation of teaching results



Final Models after Initial Analysis

- Unidimensional model
  - Correlation between Course and Instructor SET latent variables = .88
- No interaction terms
  - NEW Model 5: Model 4 + student and instructor covariates
    - Replaces:
      - Model 5 (Model 4 + interaction terms b/w course characteristics and response rate)
      - Model 6 (Model 5 + student and instructor covariates)

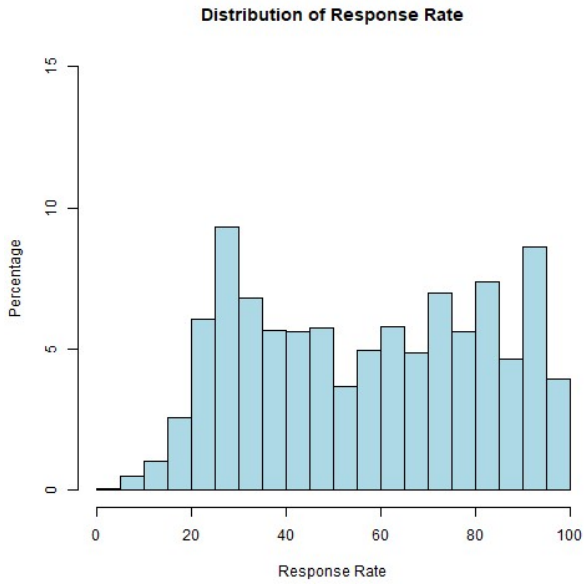
RESULTS

Data Exploration

Response Rate

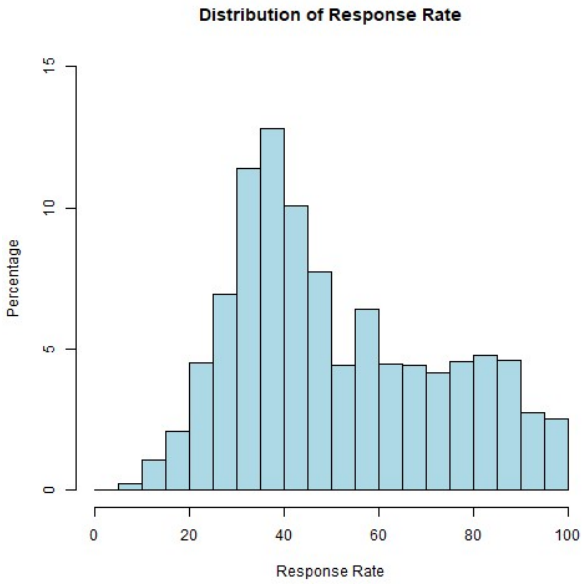
Fall 2019

- Average response rate: 57.86%



Fall 2020

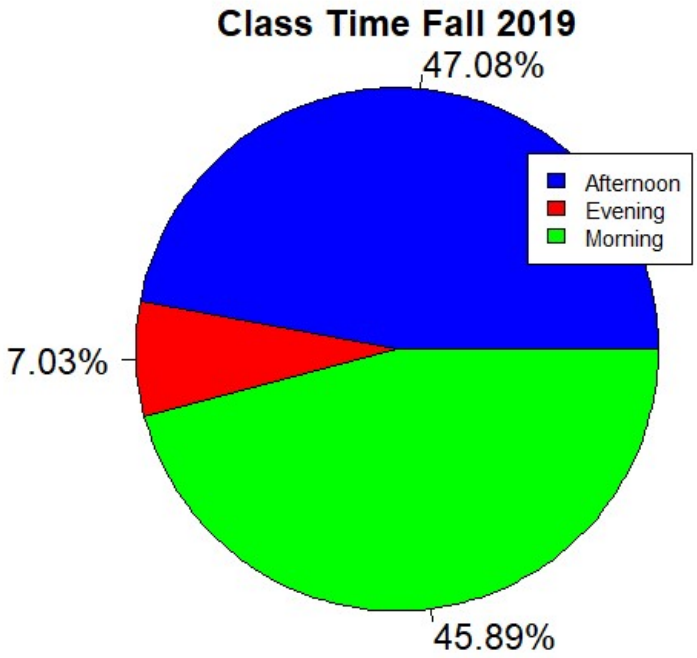
- Average response rate: 51.80%

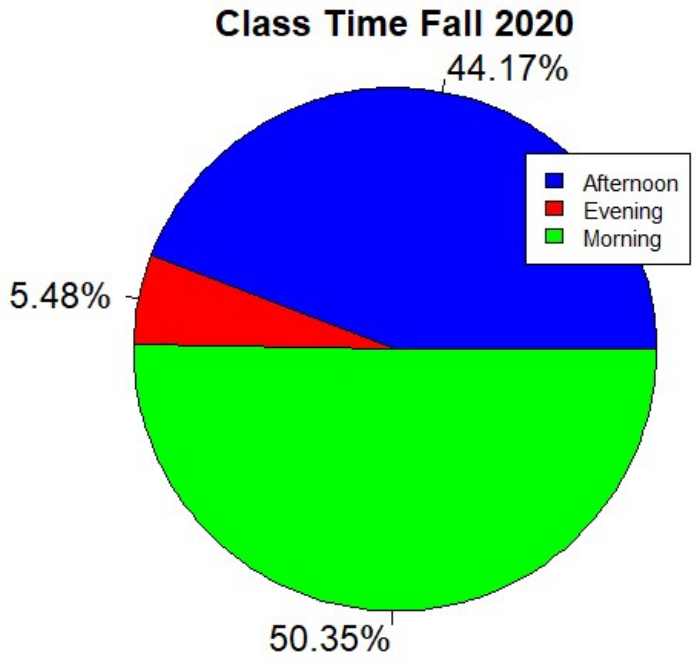


Course Characteristics

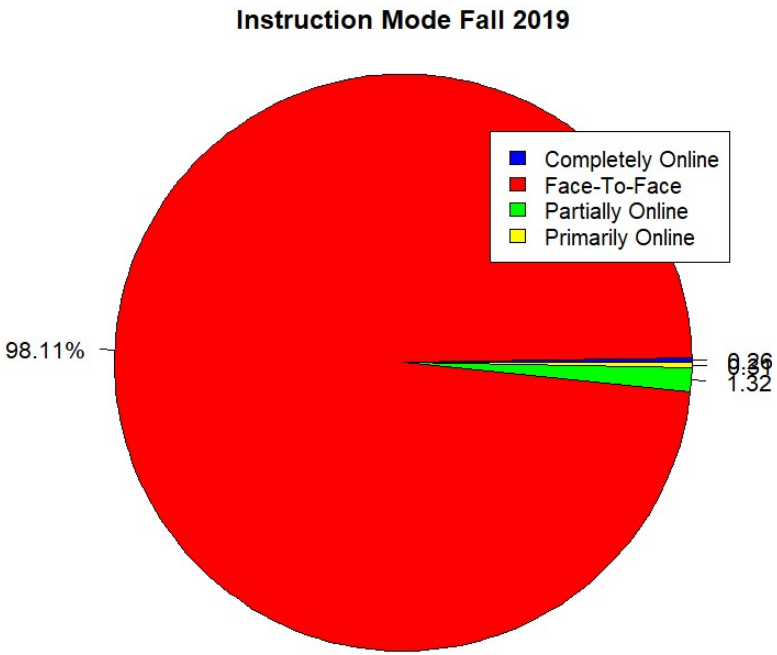
Course Characteristics	Statistic	Fall 2019	Fall 2020	Difference
Gateway	%	26.70	35.00	8.30
STEM	%	34.50	49.25	14.75
Graduate	%	12.00	9.50	-2.50
Enrollment Total	Mean	41.58	52.60	11.02
	Median	26.00	30.00	4.00
Drop/Fail/Withdrawal	%	5.38	6.03	0.65

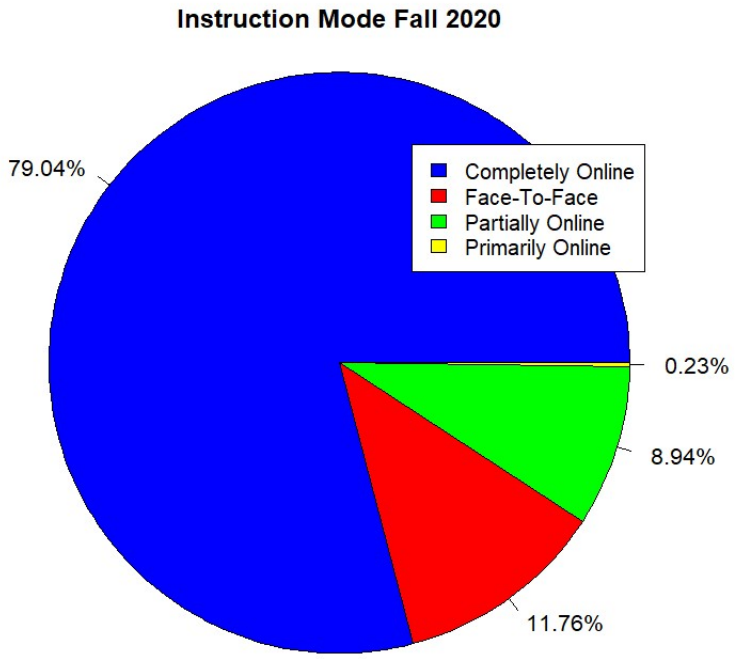
Course Characteristics: Class Time





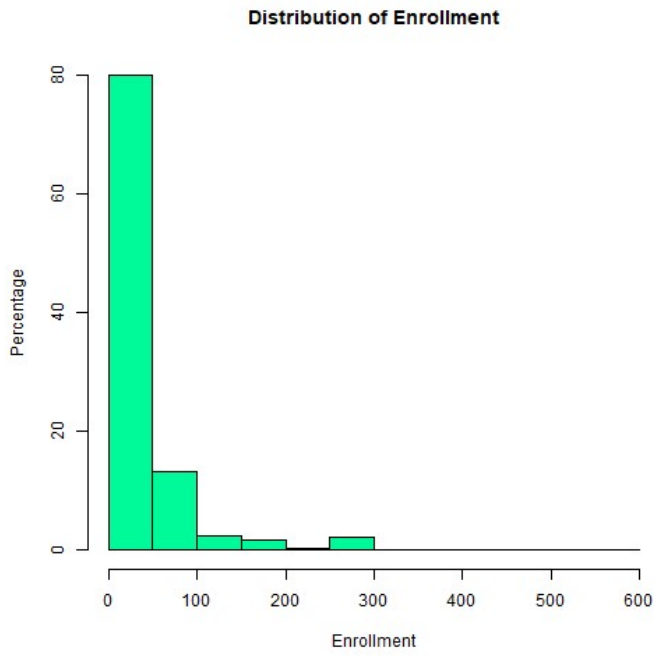
Course Characteristics: Instruction Mode



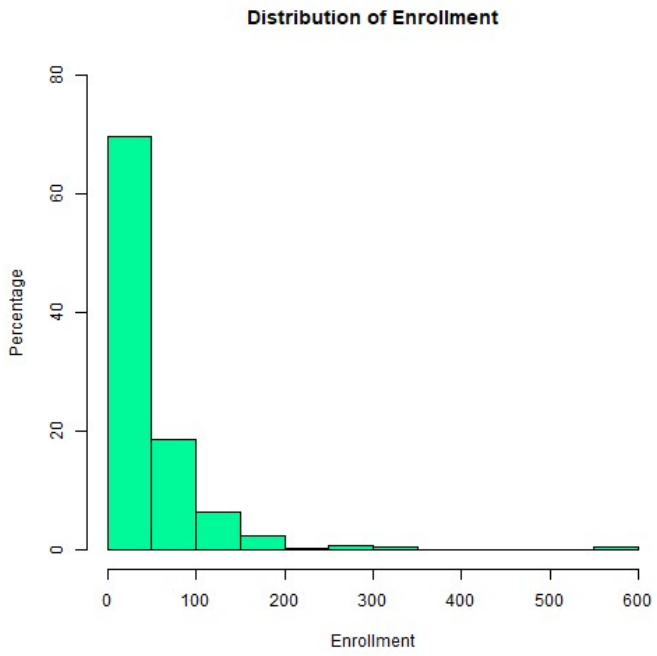


Course Characteristics: Class Size

Fall 2019

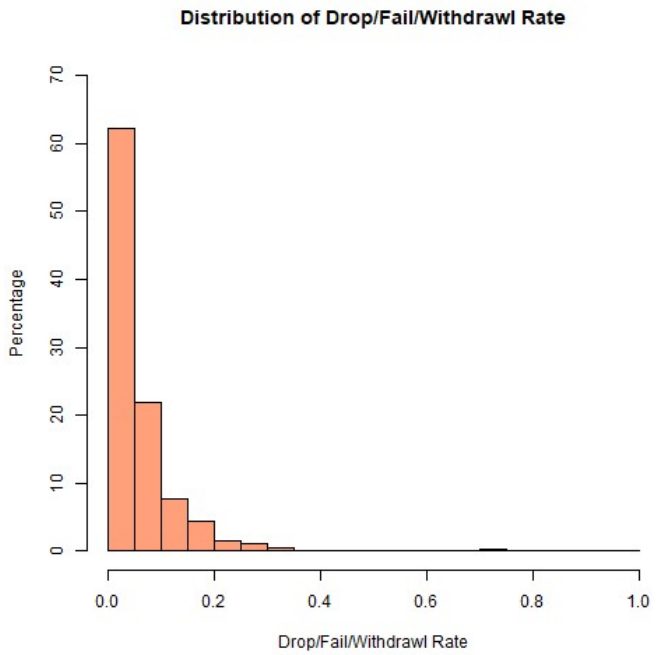


Fall 2020

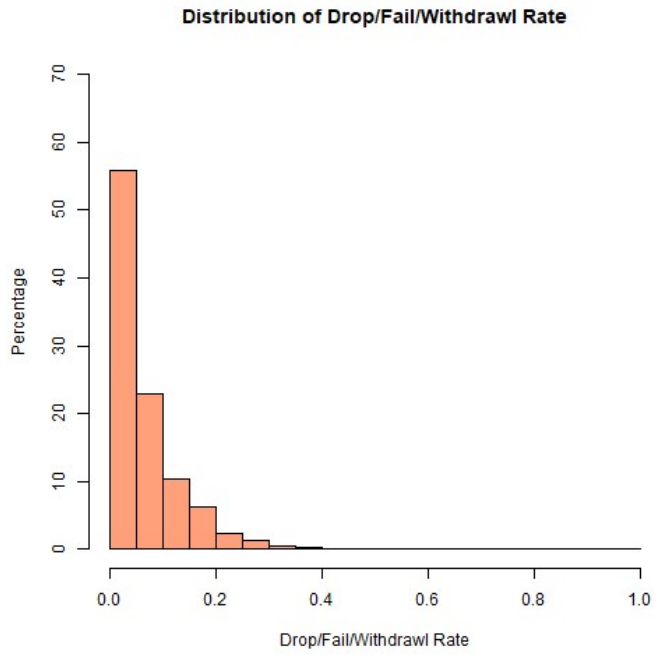


Course Characteristics: DFW Rate

Fall 2019



Fall 2020



Student Characteristics

Student Characteristics	Statistic	Fall 2019	Fall 2020	Difference
International	%	0.29	0.18	-0.11
First Generation	%	25.36	18.63	-6.73
Minorities	%	21.30	23.32	2.02
Male	%	32.78	38.68	5.90
PELL Recipients	%	20.38	14.91	-5.47
Composite ACT Score	Mean	27.00	28.50	1.50
	SD	4.31	3.91	-0.40
	Median	27	29	2.00
HS GPA	Mean	3.86	3.93	0.07
Cumulative GPA	Mean	3.45	3.51	0.06
Current GPA	Mean	3.50	3.64	0.14
Passing Grade	%	99.00	99.00	0.00

Instructor Characteristics

Instructor Characteristics	Statistic	Fall 2019	Fall 2020	Difference
Tenured	%	20.72	24.18	3.46
Teaching Years	Mean	5.80	5.14	-0.66
	Median	3.11	2.82	-0.29
Doctorates Degree	%	37.72	45.45	7.73
Males	%	48.25	52.70	4.45
Minority	%	20.26	28.58	8.32
International	%	6.22	11.92	5.70

SET Outcomes

Item		2019 Statistic		2020 Statistic		t-test		
		Mean	Median	Mean	Median	Mean difference	SE	p-value
Instructor	Prepared	5.53	6.00	5.49	6.00	-0.04	0.010	<.001
	Clear	5.29	6.00	5.23	6.00	-0.06	0.013	<.001
	Feedback	5.26	6.00	5.26	6.00	0.00	0.013	0.573
	Respect	5.64	6.00	5.64	6.00	0.00	0.009	0.906
	Recommend Instructor	5.31	6.00	5.25	6.00	-0.06	0.014	<.001
	Interactions	5.38	6.00	5.28	6.00	-0.10	0.012	<.001
Course	Understand	5.34	6.00	5.24	5.00	-0.10	0.012	<.001
	Interest	5.05	5.00	4.92	5.00	-0.13	0.015	<.001
	Technology	5.20	5.00	5.09	5.00	-0.11	0.012	<.001
	Grading	5.18	6.00	5.10	5.00	-0.08	0.014	<.001
	Recommend Course	5.13	6.00	5.03	5.00	-0.10	0.015	<.001
	Activities	5.29	6.00	5.10	5.00	-0.19	0.012	<.001
	Course Site	5.33	6.00					
	Effort			5.07	5.00			
Combined	Sum Instructor	32.42	35.00	32.15	34.00	-0.27	0.130	<.001
	Sum Course	36.53	38.00	35.54	36.00	-0.99	0.061	<.001
	Sum SRT	68.94	72.00	67.69	70.00	-1.25	0.076	<.001

RESULTS

Modeling

Results: Model 1 (Null model)

Intraclass correlation (ICC)

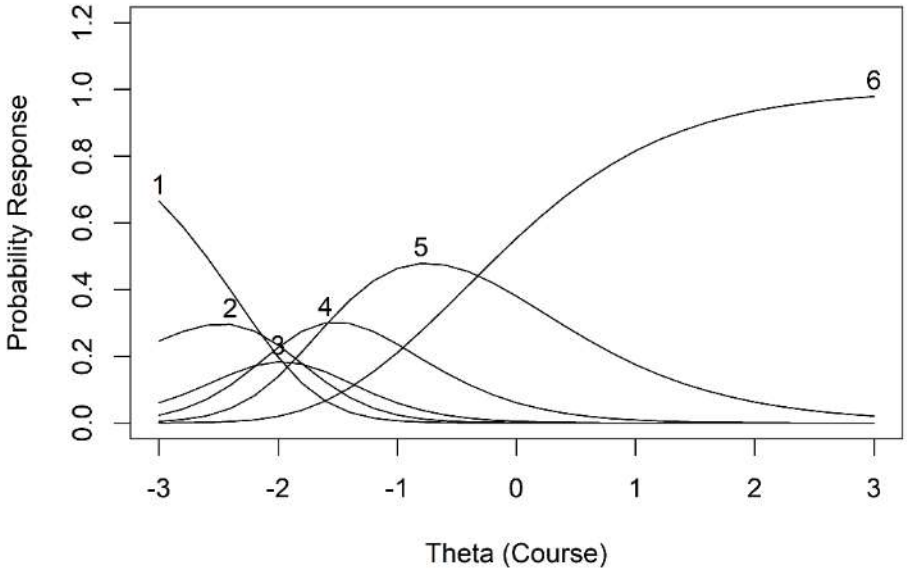
2020

- $ICC_{courses} = .21 \rightarrow$ 21% of variability is associated with courses
- $ICC_{students} = .29 \rightarrow$ 29% of variability is associated with students

2019

- $ICC_{courses} = .19$
- $ICC_{students} = .36$

Item characteristic curve (ICC)



An example of Recommend Course item with an average student

Results: Model 2 (Course characteristics)

2020

$$\hat{\xi}_g = -0.22(Gateway) - 0.30(STEM) - 0.00(Class\ size) - 1.45(DFW\ rate) + 0.01(Class\ time: Afternoon) - 0.02(Class\ time: Evening) + 0.13(Course\ level: Grad) - 0.08(Instruction\ mode: CO) + 0.01(Instruction\ mode: PA) + 0.86(Instruction\ mode: PR)$$

	Logit	Odds	Probability
Gateway	-0.22	0.80	0.45
STEM	-0.30	0.74	0.43
DFW rate	-1.45	0.23	0.19
Course level_Grad	0.13	1.14	0.53
Instruction mode_PR	0.86	2.36	0.70

Results: Model 2 (Course characteristics)

2020

$$\text{LogOdds} = \log\frac{\pi}{1-\pi} = -0.22(Gateway)$$

$$\text{Odds} = \exp(\frac{\pi}{1-\pi}) = \exp(-0.22) = 0.80$$

$$\pi = \exp(\frac{0-0.22}{1-(0-0.22)}) = 0.45$$

If the probability of responding a 6 vs. a 5 for a non-gateway course is 50%, then the probability of responding a 6 vs. a 5 for a gateway course is 45%. We see that the probability of scoring a 6 vs. a 5 is LOWER for a gateway course than a non-gateway course.

Results: Model 2 (Course characteristics)

2020

$$\text{LogOdds} = \log\frac{\pi}{1-\pi} = 0.86(Instruction\ Mode: PR)$$

$$\text{Odds} = \exp\left(\frac{\pi}{1-\pi}\right) = \exp(0.86) = 2.36$$

$$\pi = \exp\left(\frac{0+0.86}{1-(0+0.86)}\right) = 0.70$$

If the probability of responding a 6 vs. a 5 for a course taught in person is 50%, then the probability of responding a 6 vs. a 5 for a course taught primarily online is 70%. We see that the probability of scoring a 6 vs. a 5 is HIGHER for a course taught primarily online than a course taught in person.

Results: Model 2 (Course characteristics)

2020

$$\hat{\xi}_g = -0.22(Gateway) - 0.30(STEM) - 0.00(Class\ size) - 1.45(DFW\ rate) + 0.01(Class\ time: Afternoon) - 0.02(Class\ time: Evening) + 0.13(Course\ level: Grad) - 0.08(Instruction\ mode: CO) + 0.01(Instruction\ mode: PA) + 0.86(Instruction\ mode: PR)$$

2019

$$\hat{\xi}_g = -0.43(Gateway) - 0.10(STEM) - 0.00(Class\ size) + 0.00(DFW\ rate) + 0.01(Class\ time: Afternoon) - 0.19(Class\ time: Evening) - 0.10(Course\ level: Grad) - 0.28(Instruction\ mode: CO) - 0.06(Instruction\ mode: PA) - 0.06(Instruction\ mode: PR)$$

Results: Model 3 (Response rate)

2020 $\hat{\xi}_g = 0.01(Response\ rate)$

2019 $\hat{\xi}_g = 0.00(Response\ rate)$

Results: Model 4
(Course characteristics + Response rate)

2020

$$\hat{\xi}_g = -0.21(Gateway) - 0.29(STEM) - 0.00(Class\ size) - 1.37(DFW\ rate) + 0.01(Class\ time: Afternoon) - 0.01(Class\ time: Evening) + 0.13(Course\ level: Grad) - 0.07(Instruction\ mode: CO) + 0.01(Instruction\ mode: PA) + 0.86(Instruction\ mode: PR) + 0.00(Response\ rate)$$

2019

$$\hat{\xi}_g = -0.45(Gateway) - 0.10(STEM) - 0.00(Class\ size) + 0.01(DFW\ rate) + 0.01(Class\ time: Afternoon) - 0.18(Class\ time: Evening) - 0.15(Course\ level: Grad) - 0.25(Instruction\ mode: CO) - 0.07(Instruction\ mode: PA) + 0.00(Instruction\ mode: PR) + 0.00(Response\ rate)$$

Results: NEW Model 5
(Model 4+ Instructor & student covariates)

2020

- $\hat{\xi}_g = -0.25(Gateway) - 0.24(STEM) - 0.00(Class\ size)$
  $-1.19(DFW\ rate) + 0.03(Class\ time: Afternoon)$
  $- 0.01(Class\ time: Evening) + 0.13(Course\ level: Grad)$
  $- 0.07(Instruction\ mode: CO) - 0.01(Instruction\ mode: PA)$
  $+ 0.91(Instruction\ mode: PR) + 0.00(Response\ rate)$
  $- 0.12(Faculty\ tenure) - 0.01(Faculty\ time\ in\ jobs)$
  $- 0.03(Faculty\ doctorate) - 0.19(Faculty\ gender: Male)$
  $- 0.13(Faculty\ minority) - 0.12(Faculty\ International)$
- $\hat{\eta}_j = -0.10(International) + 0.00(First\ generation)$
  $+0.05(Minority) + 0.08(Male) - 0.01(PELL) - 0.00(ACT\ score)$
  $+ 0.03(HS\ GPA) + 0.15(Cur\ GPA) - 0.12(Cum\ GPA)$
  $+ 0.34(Course\ Grade: Pass) - 0.00(Unit\ taken)$

Results: NEW Model 5
(Model 4+ Instructor & student covariates)

2019

- $\hat{\xi}_g = -0.44(Gateway) - 0.11(STEM) - 0.00(Class\ size)$
  $+0.01(DFW\ rate) - 0.01(Class\ time: Afternoon)$
  $- 0.22(Class\ time: Evening) - 0.23(Course\ level: Grad)$
  $- 0.29(Instruction\ mode: CO) - 0.08(Instruction\ mode: PA)$
  $- 0.03(Instruction\ mode: PR) + 0.00(Response\ rate)$
  $+ 0.06(Faculty\ tenure) - 0.02(Faculty\ time\ in\ jobs)$
  $- 0.02(Faculty\ doctorate) + 0.03(Faculty\ gender: Male)$
  $- 0.23(Faculty\ minority) + 0.13(Faculty\ International)$
- $\hat{\eta}_j = 0.12(International) + 0.01(First\ generation)$
  $+0.01(Minority) - 0.01(Male) - 0.06(PELL) - 0.00(ACT\ score)$
  $+ 0.00(HS\ GPA) + 0.01(Cur\ GPA) - 0.04(Cum\ GPA)$
  $+ 0.01(Course\ Grade: Pass) - 0.00(Unit\ taken)$

Discussion: Summary

- An enterprise-wide understanding of how the courses themselves may influence the SET
  - Gateway courses, STEM fields, DFW rates, mode of course administration
  - Comparison of during pandemic and pre-pandemic

- Advanced modeling approach

(a) accommodates categorical variables

(b) accounts for cross-classified multilevel data structure

(c) allows for the inclusion of observed predictors and covariates in a latent factor model

Discussion

- Significant course characteristic variables are:
  - Course type (Gateway, STEM, DFW rate)
  - Course level (Graduate)
  - Instruction mode (PR)
- Response rate is not related to student evaluation of teaching.
  - Response rate does not differently affect the relationship between course characteristics and item-level responses.
- Significant instructor & student characteristic variables are:
  - Instructor: tenure status, gender, minority
  - Student: minority, gender, current GPA, cumulative GPA, course grade

Discussion

- Bias or “true” measure of teaching effectiveness?
- Some variables were not significant, but close, due only because of small sample
- Cautious interpretation on the Fall 2020 results
  - A potential impact of the COVID-19 pandemic
  - Implications for future course designs

Discussion

- Sample size
  - Missing data handling
- Interactions among course-, instructor-, and student-related variables
- Nonlinear relationship with SET
- A closer look at variables of interest
  - Focus group
  - Individual interview
  - Targeted survey
- Extension to multidimensional model

Discussion

With future research, guide department heads, chairs, and tenure review committee members on the correct use of the instructor and course evaluations

- Guidance on the usage of the course evaluation for teaching effectiveness
- Course design characteristics that can be implemented to improve student experience
- Additional support for students

Acknowledgements

- Thank you for listening.

- We would like to thank the BlueNotes Group for sponsoring this work, and Melanie Forstie (Office of Measurement Services) and Office of Institutional Research for providing data.

- Any errors are on my own.

UNIVERSITY OF MINNESOTA

Driven to Discover®

Crookston Duluth Morris Rochester Twin Cities