What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

Introduction
Motivation

Findings
Evaluation Criteria

Methods

Findings
Context
Work distribution
Pairing
Undergrad/Grad
Gender
Experience
Grades
Retention

Discussion
Open questions
Implications

What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

University of Toronto & University of British Columbia
WCCCE 2012

May 4, 2012

Why study TAs?

What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

Introduction
Motivation

Findings
Evaluation Criteria

Methods

Findings
Context
Work distribution
Pairing
Undergrad/Grad
Gender
Experience
Grades
Retention

Discussion
Open questions
Implications

- ▶ They improve student success in CS1 (Wilson, 2001), especially for minorities (Roberts, 1995)

- ▶ TAs give 46% of our contact hours in 1st & 2nd-year CS

- ▶ Issues of TA quality

Why TA evaluations?

What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

Introduction
Motivation

Findings
Evaluation Criteria

Methods

Findings
Context
Work distribution
Pairing
Undergrad/Grad
Gender
Experience
Grades
Retention

Discussion
Open questions
Implications

▶ Evaluations are an imperfect measure of teaching

▶ Nevertheless, good TA evaluations encourage TAs (Bomotti, 1994)

▶ TAs are new teachers, flexible in their approach to teaching (Muzaka, 2009)
  ▶ They are also hungry for feedback (Patitsas, 2012)

Studying TAs

What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

Introduction
Motivation

Findings
Evaluation Criteria

Methods

Findings
Context
Work distribution
Pairing
Undergrad/Grad
Gender
Experience
Grades
Retention

Discussion
Open questions
Implications

▶ TAs are linked to:
  ▶ student retention (O'Neal, 2007)
  ▶ student grades (Paul, 2010)
  ▶ and provide role models to our students (Patitsas, 2012)

▶ In this talk, we will be examining a number of such matters; however, we'll be leaving more questions than answers.

We ask...

What would you like to know about your TAs?

What do you think your TAs should know about their teaching?

Evaluation Criteria

The five criteria are evaluated on a 5-point Likert scale, and are:

- ▶ Well prepared (W.P.)

- ▶ Helpful (Help.)

- ▶ Considerate of students (Consid.)

- ▶ Easily understood (E.U.)

- ▶ An effective instructor (E.I.)

Delving into findings...

What can we learn
from quantitative
teaching assistant
evaluations?

Elizabeth Patitsas
and Patrice
Belleville

▶ We acquired 231 anonymized TA evaluations

▶ And we looked at the correlations between those criteria

▶ Pop quiz: what is the correlation coefficient needed to
  be a "relationship" in the social sciences?

Evaluation Criteria, contd.

What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

Introduction
Motivation

Findings
Evaluation Criteria

Methods

Findings
Context
Work distribution
Pairing
Undergrad/Grad
Gender
Experience
Grades
Retention

Discussion
Open questions
Implications

	W.P.	Help.	Consid.	E.U.	E.I.
Well prepared					
Helpful	0.8				
Considerate	0.7	0.8			
Easily understood	0.7	0.8	0.6		
Effective instr.	0.8	0.9	0.8	0.9	

- All are $p < 0.001$
- Coarse-grained evaluations!

Methods

1. Literature review: what variables are worth testing?

2. Look at those variables

3. Threshold for significance: $p < 0.001$

4. n=231, so large sample size

So what did we learn?

Context: CS at UBC

- First and second year courses at UBC have:
  - 3 hours of lecture / week
  - One 2 or 3 hour lab session / week
  - In some cases, a 1 hour tutorial

- Lab sections:
  - Have been 20 and 25 students

- While all TAs work roughly the same number of hours, work distribution varies

Work distribution

What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

Introduction
Motivation

Findings
Evaluation Criteria

Methods

Findings
Context
Work distribution
Pairing
Undergrad/Grad
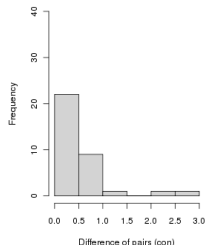Gender
Experience
Grades
Retention

Discussion
Open questions
Implications

Negative correlation: between how many lab sections a TA taught and their TA evaluations (r=-0.4).

  ▶ TAs with 1-2 lab sections had better evaluations than TAs with 3
  ▶ TAs with 4 lab sections a week performed worst of all

Open question: why? Boredom effect? Contact hours are more tiring?

Pairing
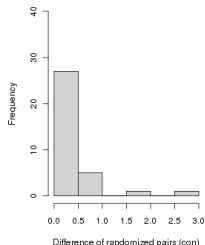
We found that paired TAs received similar evaluations.

Open question: *Do better TAs bring their partners up? Or do students just rank pairs as a unit? Is the lab only as good as how well the TAs work together?*

Context: Types of TAs

What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

Introduction
Motivation

Findings
Evaluation Criteria

Methods

Findings
Context
Work distribution
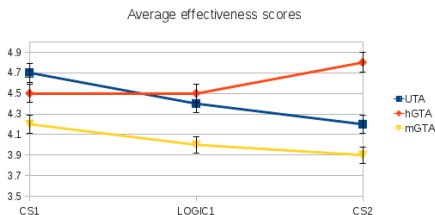Pairing
Undergrad/Grad
Gender
Experience
Grades
Retention

Discussion
Open questions
Implications

- ▶ Graduate Teaching Assistants (GTAs):
  - ▶ Monthly GTAs
    - ▶ The TA work is part of their promised funding.
  - ▶ Hourly GTAs
    - ▶ They apply for the position, we select the hourly GTAs.
- ▶ Undergraduate Teaching Assistants (UTAs):
  - ▶ They apply for the position, select the UTAs we hire.

Undergraduate vs. Graduate TAs

What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

Introduction
Motivation

Findings
Evaluation Criteria

Methods

Findings
Context
Work distribution
Pairing
Undergrad/Grad
Gender
Experience
Grades
Retention

Discussion
Open questions
Implications

Average effectiveness scores

That UTAs performed better in teaching first-year has been found at other institutions (Mendenhall, 1983)

Open question: *Why? Course-specific knowledge vs. conceptual knowledge?*

Gender

Previous work found biases against female instructors in evaluations (Wachtel, 1998)

We found no overall differences between how female and male TAs were ranked by their students.

Experience

No statistically significant link between course-specific experience and TA evaluations.

Threats to validity: we could only match TAs within a course, and only within the time period we sampled.

Experience, contd.

While there is evidence that TAs improve as teachers over time, the effect of their experience may be counterbalanced by more "sterner" teaching (Patitsas, 2012).

Previous work on teaching evaluations has found no link between experience and evaluations (Wachtel, 1998).

Students' Grades

The literature finds that TAs affect students' performance (Paul, 2010)

We found no link between students' final grades and their TAs' evaluations.

This does not discount the likelihood that a TA has an effect on their students' performance

Whatever effect a TA has on their students is *not* captured by these quantitative evaluations.

What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

Introduction
Motivation

Findings
Evaluation Criteria

Methods

Findings
Context
Work distribution
Pairing
Undergrad/Grad
Gender
Experience
Grades
Retention

Discussion
Open questions
Implications

Programme-Level Retention

What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

Introduction
Motivation

Findings
Evaluation Criteria

Methods

Findings
Context
Work distribution
Pairing
Undergrad/Grad
Gender
Experience
Grades
Retention

Discussion
Open questions
Implications

Previous literature has found that a negative lab atmosphere
will contribute to students' decision to take
fewer science courses (O'Neal, 2007).

We found a weak, correlation between low-ranked TAs'
evaluations and how many more computer
science classes their students took.

Open question: *are TAs scaring away students, or are
unengaged students are rating TAs down?*

Open questions

How can we provide TAs with effective formative assessment?

To what extent are quantitative student evaluations useful? What questions would be better?

How could qualitative evaluations be better harnessed?

Why are UTAs ranked better than GTAs?

Causality: student retention, TA evals

Causality: pairing

Implications and Suggestions

What can we learn from quantitative teaching assistant evaluations?

Elizabeth Patitsas and Patrice Belleville

Introduction
Motivation

Findings
Evaluation Criteria

Methods

Findings
Context
Work distribution
Pairing
Undergrad/Grad
Gender
Experience
Grades
Retention

Discussion
Open questions
Implications

▶ TA evals are coarse-grained, need more info to make award/hiring decisions

▶ Probably not very useful to the TAs; part of the TA quality issue?

▶ Put UTAs on first-year

▶ Keep TA workloads reasonable

▶ Pair your TAs

Acknowledgments