Predicting Teaching Evaluation Scores and the Importance of Multimode Measurement

Center for Teaching Excellence
Loyola Marymount University
17 January 2013
• **Committee on Comprehensive Evaluation of Teaching**
  - Charged to recommend ways of evaluating teaching
  - Invited IR to hear hypotheses about evaluation scores
  - Commissioned IR to produce a report on any evidence

• **Institutional Research**
  - Generates campus’ official statistics
  - Conducts original research for campus decision making
Method: Data Overview

- Fall 2009 to spring 2011 evaluation ballots
  - Entire history of new teaching evaluation form
  - “Overall Instructor Effectiveness” item used as outcome

- Anonymous ballots imply course-level analysis
  - Avoid ecological fallacy
  - Results say nothing about tendencies of individual students

- Undergraduate courses only
  - Cross lists combined, coded as section with most enrollment
  - Team-taught courses coded for primary instructor only
Method: More Data

• **Banner database for course-level characteristics**
  - Grades, enrollment, time, location
  - New preparation (not taught in last 5 years)
  - Percent of students in class
    - Of the subject’s major
    - Male or female

• **Human Resources database for instructor info**
  - Birthday, gender, and ethnicity
  - Rank (5 buckets): PT, Visiting/Clinical, Asst., Asso., Full

• **RateMyProfessors.com (worth the 14% loss of sample)**
Method: Model

- Ordinary Least Squares regression
- Explore for general predictors
- Test CCET hypotheses using predictors as controls
  - Do ratings differ by instructor rank, gender, ethnicity?
  - Are there differences by college?
  - What about courses with high proportions of students:
    - Fulfilling an LMU core requirement?
    - Fulfilling a major requirement?
    - Who are majors in that subject?
## Results: General Predictors

Six relationships explain 33% of ratings variation

<table>
<thead>
<tr>
<th>Factor</th>
<th>Coef.</th>
<th>$t$</th>
<th>$p$</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Student Interest</td>
<td>0.39</td>
<td>17.22</td>
<td>0.00</td>
<td>4 X</td>
</tr>
<tr>
<td>Chili Pepper</td>
<td>0.20</td>
<td>6.49</td>
<td>0.00</td>
<td>2 X</td>
</tr>
<tr>
<td>Instructor Age</td>
<td>-0.01</td>
<td>-5.84</td>
<td>0.00</td>
<td>2 X</td>
</tr>
<tr>
<td>New Preparation</td>
<td>-0.17</td>
<td>-5.43</td>
<td>0.00</td>
<td>1 X</td>
</tr>
<tr>
<td>Avg. Grade in Class</td>
<td>0.16</td>
<td>5.31</td>
<td>0.00</td>
<td>1 X</td>
</tr>
<tr>
<td>Class Duration (hours)</td>
<td>-0.07</td>
<td>-5.15</td>
<td>0.00</td>
<td>1 X</td>
</tr>
</tbody>
</table>

Ensure apples-to-apples comparisons in subsequent analyses
Results: CCET Questions

- **Instructor gender** – No change
  - After accounting for age, positive female effect vanished
  - Literature suggests effects in sciences, but not at LMU
- **Instructor rank** – Associates \( \rightarrow \) -0.1 pt. vs. assistants
- **Course college** – SE courses \( \rightarrow \) -0.1 pt. vs. BCLA
- **For core fulfillment** – No core to all \( \rightarrow \) +0.2 pt.
- **For major requirement** – No difference
- **Majors in course** – None to all \( \rightarrow \) -0.1 pt.
Results: Ethnicity

Among courses taught by all faculty:

- Asians 0.24 pt. less on average
- Black/African-Am. 0.29 pt. less

Among courses taught by tenure line faculty:

- Asians 0.18 pt. less on average
- Black/African-Am. 0.34 pt. less
• General negative effect for faculty of color
  - Female minority vs. white ➔ -0.5σ (Hamermesh & Parker, 2005)
  - Black/African-Am. vs. white ➔ -0.5σ (Smith, 2007)
  - Non-white vs. white ➔ -0.3σ (McPherson et al., 2009)
  - Web ratings / race perceptions with public faculty info (Reid, 2010)
    • Asian vs. white/Hispanic ➔ -0.1σ to -0.2σ in multiple measures
    • Black/African-Am. ➔ -0.4σ in multiple measures

• Reasons hypothesized
  - Student centered
    • Preconceptions (Smith, 2007; Anderson & Smith, 2005; Chowdhary, 1988)
    • Prejudices on ethnicity and subject competency (Hendrix, 1998)
  - Language (Hamermesh & Parker, 2005; Rubin, 1998; Rao, 1995)
- Adjust ratings for factors (McPherson et al., 2009)
- Dialog (Chamberlin & Hickey, 2001)
  - Teaching styles and their relationship to ratings
  - Use of ratings in merit and promotion
- Include measurements of teaching technique
  - Feldman, 1993
  - Ludwig & Meachum, 1997
  - Crombie et al., 2003)
Committee on the Comprehensive Evaluation of Teaching (CCET)

17 January 2013
Center for Teaching Excellence
Comprehensive Evaluation of Teaching

Goals of Teaching Evaluation
• Data Collection and Assessment
• Feedback for Improvement

Ideal Features of Teaching Evaluation
• Relevant
• Reliable ("Discriminating")
• Consistent
• Objective
• Systematic
## Evaluating Relevant Aspects of Teaching

<table>
<thead>
<tr>
<th></th>
<th>Self</th>
<th>Students</th>
<th>Peer Observation</th>
<th>Chair/Dean</th>
<th>R&amp;T Committee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Expertise</td>
<td>Yes</td>
<td>No</td>
<td>Usually</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Class Design and Organization</td>
<td>Yes</td>
<td>Somewhat</td>
<td>Yes</td>
<td>Maybe</td>
<td></td>
</tr>
<tr>
<td>Presentation Skills</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Rapport with Students</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Course Management</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Student Achievement</td>
<td>Yes</td>
<td>Somewhat</td>
<td>No</td>
<td>Somewhat</td>
<td></td>
</tr>
</tbody>
</table>

This table provides a generic picture only. Here, a Chair or Dean may or may not be in the same field; they base their evaluations on available documentation, including syllabi, assignments, administrative information, etc.

### Ability to evaluate depends on
- **Clear standards.**
- **Quality and quantity of info available to each party.**
- **Experience, training, and personality of evaluator.**
- **Comprehensive Evaluation of Teaching**
- **Clear Standards**
- **Support and Training**
Evaluating Teaching

Comprehensive Evaluation of Teaching

The Faculty Senate Committee on the Comprehensive Evaluation of Teaching (CCET) was charged in the spring of 2010 "to investigate, document, and assist with the development of tools for evaluating teaching beyond student evaluations. The committee will analyze different options, develop appropriate guidelines, and make recommendations to the Faculty Senate regarding more comprehensive methods of evaluating teaching at LMU."

The committee’s work, as charged by the Senate, is based on the principles that

1. A comprehensive evaluation of teaching will require multiple measures, because effective teaching involves multiple dimensions and any particular instrument has definite limitations. Attempting to evaluate teaching ability with a single measure does a disservice both to the university and the faculty member.

2. The evaluation of teaching should take place for two purposes: (a) to provide information used to make decisions on retention and merit (i.e., for FSRs) and for applications for advancement to tenure or in rank (summative evaluation) and (b) to provide information that could be used to improve teaching at every stage of a faculty member’s career (formative evaluation).

The Teaching Toolkit will provide guidelines, supporting documents, and implementation/application suggestions for all parties involved in the evaluation of teaching. The Teaching Toolkit is currently in development. All documents made available here are currently in draft form. Feedback is welcome!

See http://www.lmu.edu/libraries_research/cte/Resources/Evaluating_Teaching.htm