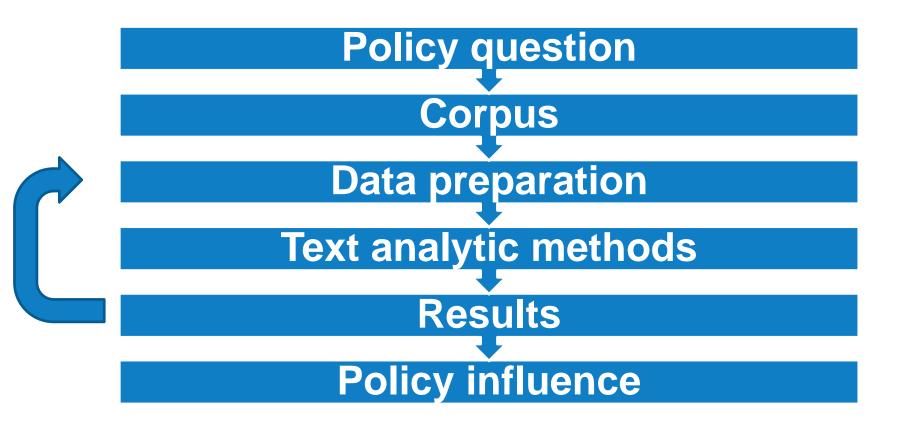
Using Text Analysis of "Most Meaningful Experiences" to Inform Policy

Pamela Brown, Joseph Van Matre, and Zachary Bleemer

Agenda

- Introduction
- Framework
- Policy Question
- Text Analysis Strategies
 - Word use frequency
 - Sentiment analysis
 - Topic models
- Policy influence

Analytical framework



NOTE: Sample R code will be made available for all methods described in this presentation

Policy Question: Strategic Planning for UC 2030

University of California aims to improve graduation rates by 2030

COMMENTARIES

Focus is on low-income, African-American, Latino and first generation students.

PROJECTS



HIGHLIGHTING STRATEGIES FOR STUDENT SUCCESS

SEPTEMBER 20, 2019





TOPICS

UC 2030 Dashboard

AUDIO & VIDEO

UC has goals to (1) produce 200,000 more undergraduate and graduate degrees on top of the 1 million already projected; (2)achieve a 90 percent overall graduation rate and eliminate gaps for Pell, first-generation and underropresented groups; and (3) invest in faculty and research by adding 1,100 ladder rank faculty over the next four years. This dashboard presents systemwide and campus goals and a means to track progress. State funding is a critical component to success and UC has requested 560 million in permanent funding over the next four years. UC will present any funding received for the system with allocations by campus.

PUBLICATIONS

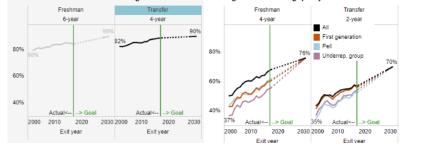
ABOUT EDSOURCE

Award 1.2 million degrees between 2015-16 and 2029-30

DATA



Increase freshman and transfer graduation rates Close graduation rate gaps by 2030



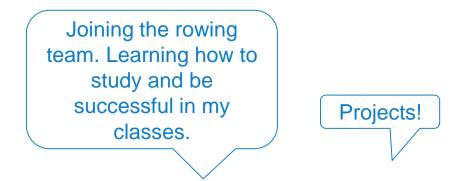
Select Campus

niversitywide

Define your corpus

What is one of the most meaningful learning experiences you have had at "your UC campus"?

Taking sociology with [professor]. Hearing where he came from and the experiences he went through to be able to make it to this level of academia was very inspiring.



Learning you won't be reached out to, you have to be the one that reaches out.



UNIVERSITY OF CALIFORNIA

Data Preparation:

- Merge survey responses to student characteristics
- Generate list of all popular words in responses
 - Omit words that appear fewer than 50 times
 - Omit "stop words" like prepositions
 - "De-stem" words (e.g. remove plural 's' and verb conjugations)
 - Remove "invalid" words that will impair results (e.g. campus names)
- Add common two- or three-word phrases

Three text analysis tools:

1. Word use frequency by group

2. Sentiment Analysis

3. Topic Models

Methodology:

- Produce a data frame with:
 - Student characteristics, usually defined as binary variables
 - 0/1 indicators for every identified word
- Estimate linear regressions of each student characteristic on all of the word indicators
- Produce word clouds of the words with the highest estimated t-statistics (from the null hypothesis)
 - t-statistics are higher if (1) the word is more common (bypassing outliers) and (2) the word disproportionately appears in the responses of students with the given characteristic

Word frequency examples: simple word cloud

meaningful learning experiences learning experiences school thing also ableresearch world time professor different way people student taught community made work life learn real opportunity year **EXPERIENCE** major lot **Learning** friend study meaningful learning like course campus learning experience

Word frequency examples: UC campuses

UC Davis

UC Santa Cruz

UC Merced

connecting experience freshmen understood childadvisorleading connection conference community design internship useful animal effect useful animal studying degree native clinic handson bad feel clinic handson bad feel clinic handson christian intern first year expectation guarter system belong withinextreme leadership skills progress university psych travel natural model housing computer incredibly college game without advisor COICE confident white COICE quarter forever senior theater stem field org art small section org active important spacetutoring creating color intro injustice color intro meaningful experiences whether writing stay grant one another talent type add attending uc generation friend bridgepublic health journey club core strive stuck Semester study habits stuck Semester matter undocumented passing engineering leadership bit staff status campus opportunity everyone serviceassist one meaningful amazing people meaningful experience develop

Word frequency examples: Ethnicity

Asian

Hispanic

White

student organization interacting meet met memorable broaden handson computer science fight game pressure drive drive grew biothus driven iapanese growth learn aslan growth improve aslan officer real life meeting dancerigor assistant comfort feel like mature definitely business collegementorship graduate student perspective interpersonal

leadership skills knowinginjustice united states advisor meaningful experience something home navigate program latino spanish biological easierlowincome scholar support _ city part tool job feel overall manner knew able assignment help us hispanic meant meaningful experiences english. undocumented amazing open minded

best learning core two surrounding class taught survey living credit spent field fantastic far senior white physics sort rk explain coursewo man graduate thesis animal global researchsource roommate humanity Incredible acceptance incredibly literature independent research gettingfaculty requirement

Three text analysis tools:

1. Word use frequency by group

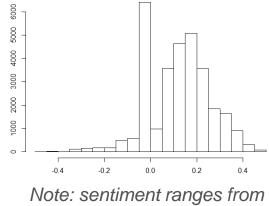
2. Sentiment Analysis

3. Topic Models

Methodology:

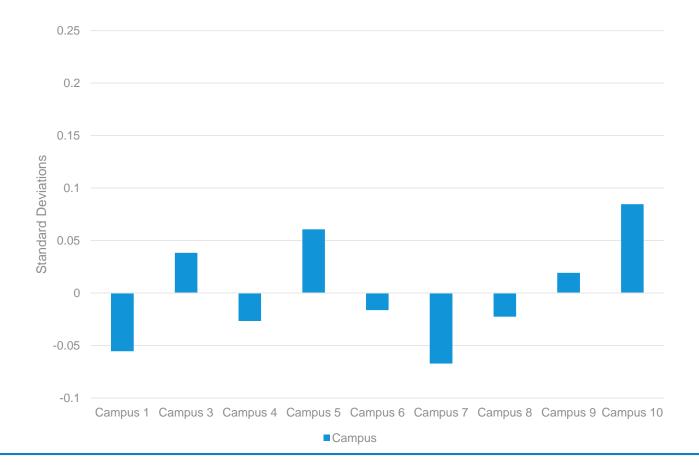
- Estimate 'Positive' and 'Negative' sentiment of each response
 - Uses the "SentimentAnalysis" R package
- Estimate linear regressions of response sentiment against student characteristics
- Produce word clouds of the words most-associated to positive and negative responses

Histogram of response sentiment measures

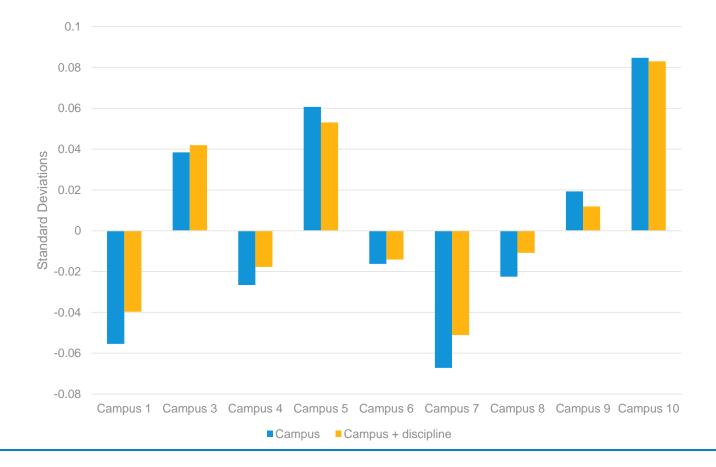


-1 to 1

Regression estimates: effect size of estimates



Regression estimates: effect size of estimates



Three text analysis tools:

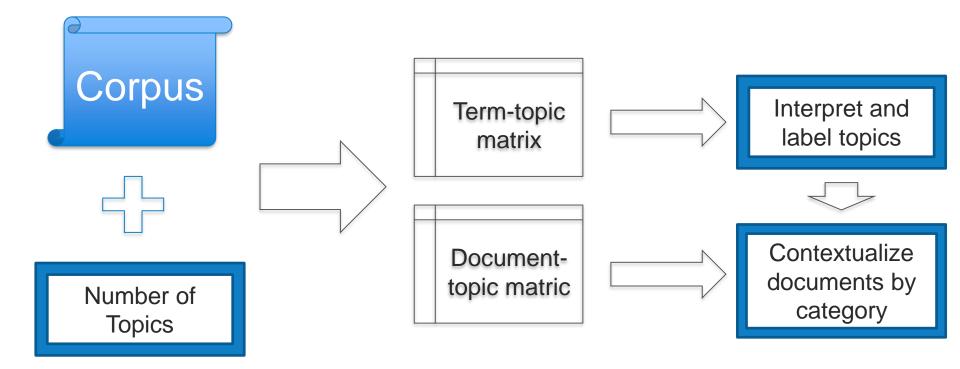
- 1. Word use frequency by group
- 2. Sentiment Analysis

3. Topic Models

Methodology:

- Computationally identify "topics" in responses
 - Uses the "topicmodels" R package
 - You choose the number of topics (we chose 30). The computer finds them, characterized by often-adjacent words
- Each topic is characterized by the words that appear more-frequently in that topic; each response is characterized by its topics
- Regress student characteristics on topic composition to identify which topic is mostassociated

Methodology:



UC Berkeley is most closely associated with topic 17, which looks like this: **UC Merced** is most closely associated with topic 10, which looks like this:

majority exist interview remember quarter system assist project setting term academically engagement knew god undocumented god undocumented interest town racist topic interest town racist topic fully relationship word debate outside write additional cultural motivate speaker aware world classroom racism application meeting complete recognize aspect

accomplishextreme academically benefitinternational family piece pretty music history sitting difficulty believe overall hardest follow deep law midterm ever word midterm want incredible exist facing today final play whatever shape final play whatever advisor essential abroad ethnicity government studied roommate understand lecturer

Example: **Pell Students** are most closely associated with topic 25, which looks like this:

Topic 25:

mostly found design closermindset occurred improving survey blackShapetouch treat belong reasontaken serve create negative attitude finding stand saw walk seek felt please imagine home cool will now voice applicable knew communication advocate invaluable positive campus

Using results to influence policy

Using the data to inform 2030 strategic goal setting:

- Identify experiences that matter
- Identify programs and strategies that have impact
- Find artifacts that tell a *representative* story

Using results to influence policy

Survey responses that highlight program impact

"For me the most meaningful learning experience was being disqualified from the university because of low grades that were caused by difficulty with anxiety that I developed at college.

However the support of my CAPS counselors, Marshall academic advisors and some faculty helped me return and succeed. These faculty that helped me...really helped me overcome my condition and regain confidence in myself."

Survey responses that support sense of belonging

"Being able to make new friends knowing you're not the only one that may experience certain struggles. It made my first year college experience somewhat more easy to handle."

"The most meaningful learning experience that I have had so far was discovering the importance of self care... I have had to deal with the most stress and anxiety I have ever had...but there are campus resources to help me deal with it." Fixed-effect word regression, sentiment analysis, and topic modeling can be implemented "out-of-the-box."

Computational text analysis is:

- Feasible
- Illuminating
- Actionable

Thank you!

pamela.brown@ucop.edu joseph.vanmatre@ucop.edu zachary.bleemer@ucop.edu

UNIVERSITY OF CALIFORNIA