Value of Preserving and Disseminating Student Research Through Institutional Repositories

Adriana Popescu
Radu Popescu
California Polytechnic State University, San Luis Obispo
The Study

Objective

• Show statistical evidence of a relation between undergraduate student research exposure in an open IR and the impact of faculty research

Caveat

• It is not about assessing IR platforms or faculty productivity or scholarly impact assessment tools
Study origin

- What is the role of IRs?
- Managing and preserving institutional community assets
- Senior projects, electronic theses, and dissertations
The question

Does an active, open IR centered on student work have a positive impact on faculty’s research impact, independent of faculty’s participation in the IR?
About Cal Poly

LEARN BY DOING
The Meaning Behind Cal Poly's Learn by Doing Signature
The Living Library Exhibit

Artist Sketch

Sketch of mustang sculpture in proposed location drawn by ARCE student, Kevin Church (Figure 22).

Figure 23: Artist Sketch by Kevin Church
Figure 5: Elevation View of Center of Geometry for Wind Loading
Undergraduate research

- Personal and professional gains for students
- Positive impact on student retention
- Faculty benefit

No studies have been identified that attempt to demonstrate a direct influence of undergraduate research activity on faculty research impact.
Methodology

- **Objective**: Investigate if student activity in IR has any effect on faculty research output

- **Data sources**:
  - Undergraduate student projects deposited in an open IR
  - Faculty publications indexed in WoS

- **Indicators**:
  - Student activity in the IR
  - Faculty research impact

- **Analysis**: Regression analysis to assess the significance of the correlation between the two sets of indicators
Methodology

**Data set 1**
- Faculty publications from six departments in the College of Science and Mathematics (CSM)

**Data set 2**
- Senior projects from the same six CSM departments deposited in DC since 2008

Impact indicators drawn from each sample are regressed against each other to investigate if there is any significant correlation
Data set 1

1. Select departments
2. Identify all authors affiliated with Cal Poly with all variations of their names (WoS)
3. For each of the six departments
   a. Obtain list of current full time faculty (as of May 2017)
   b. For each faculty, identify all variations of their name based on data obtained from Step 1
   c. Download data from InCites for all papers selected for: all names identified at 2.a + Cal Poly affiliation + year ≥ 2008
   d. Eliminate duplicate papers (two or more authors from the same department)

Result: 871 publications authored by CSM active faculty under Cal Poly affiliation between 2008 and 2017
Defining the indicators of faculty research impact

Factors affecting raw citation counts (TC)

- Field of research
- Journal Impact Factor
- Accessibility
- Language
- Type of publication
- Publication year

Journal Normalized Citation Impact (JNCI)
(Source: InCites, Clarivate Analytics)
Indicators of faculty research impact

JNCI calculated for each paper

JNCI > 1 → cited more than average
JNCI < 1 → cited less than average

JNCI_{av} - Average normalized citation count from all papers published by the faculty in each department

ΣJNCI - Total count of normalized citations from all papers published by the faculty in each department
## Data set 1 - Indicators of faculty research impact

<table>
<thead>
<tr>
<th>Article Title</th>
<th>Source</th>
<th>Publ. Year</th>
<th>Times Cited (TC)</th>
<th>Journal Expected Citations (JEC)</th>
<th>Journal Normalized Citation Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variation in the heat</td>
<td>JOURNAL OF EXPERIMENT</td>
<td>2010</td>
<td>133</td>
<td>22.11</td>
<td>6.02</td>
</tr>
<tr>
<td>Genome size is a</td>
<td>NEW PHYTOLOGIST</td>
<td>2008</td>
<td>122</td>
<td>42.84</td>
<td>2.85</td>
</tr>
<tr>
<td>The proteomic</td>
<td>JOURNAL OF EXPERIMENT</td>
<td>2010</td>
<td>111</td>
<td>22.11</td>
<td>5.02</td>
</tr>
<tr>
<td>The evolution of</td>
<td>NEW PHYTOLOGIST</td>
<td>2010</td>
<td>102</td>
<td>74.71</td>
<td>1.37</td>
</tr>
<tr>
<td>Proteomic response</td>
<td>JOURNAL OF EXPERIMENT</td>
<td>2011</td>
<td>88</td>
<td>16.98</td>
<td>5.18</td>
</tr>
<tr>
<td>The Importance of</td>
<td>PHYSIOLOGICAL AND BIOG</td>
<td>2008</td>
<td>70</td>
<td>19.45</td>
<td>3.6</td>
</tr>
<tr>
<td>FLOWERING LOCUS C...</td>
<td>GENESIS</td>
<td>2010</td>
<td>1</td>
<td>16.43</td>
<td>0.06</td>
</tr>
<tr>
<td>Effect of host lizard</td>
<td>EXPERIMENTAL AND ...</td>
<td>2013</td>
<td>1</td>
<td>5.82</td>
<td>0.17</td>
</tr>
<tr>
<td>Pyro printing: novel</td>
<td>FASEB JOURNAL</td>
<td>2014</td>
<td>1</td>
<td>0.03</td>
<td>38.44</td>
</tr>
<tr>
<td>Exploring behavioral</td>
<td>BEHAVIOUR</td>
<td>2015</td>
<td>1</td>
<td>1.53</td>
<td>0.65</td>
</tr>
<tr>
<td>Advances in...</td>
<td>MEDITERRANEAN SEA ...</td>
<td>2014</td>
<td>1</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Fisheries and ...</td>
<td>ECOSPHERE</td>
<td>2015</td>
<td>1</td>
<td>2.02</td>
<td>0.49</td>
</tr>
<tr>
<td>Sexual Selection ...</td>
<td>ETHOLOGY</td>
<td>2015</td>
<td>1</td>
<td>1.37</td>
<td>0.73</td>
</tr>
<tr>
<td>Long-term, high ...</td>
<td>SCIENTIFIC DATA</td>
<td>2016</td>
<td>0</td>
<td>0.34</td>
<td>0</td>
</tr>
<tr>
<td>Time course of lead ...</td>
<td>JOURNAL OF ...</td>
<td>2017</td>
<td>0</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**InCites data for each department**

\[ \text{JNCI} = \frac{\text{TC}}{\text{JEC}} \]

- **JEC** – very small
  - Bias in JNCI
  - Remove from data set

- **JEC** – not available
  - Remove from data set

662 papers remaining

**Used in analysis**
Data set 2

- Senior projects deposited in Digital Commons (DC) for each of the six departments analyzed
- Downloads for projects deposited in DC for each of the six departments
- Source: DC dashboard
Data set 2 - Indicators of student research impact

Undergraduate research IR activity

DC Dashboard

Project counts (Sp): quantitative measure of student research activity normalized by number of faculty in each department Sp/NF

Project download counts (Sd): quantitative measure/qualitative measure of student research activity normalized by number of faculty in each department Sd/NF
Regression analysis explores the relation between two sets of measurements, termed as dependent variable and independent variable.

**Dependent variable**
- Research impact of faculty publications
  - JNCIav
  - ΣJNCI

**Independent variable**
- Undergraduate research activity in IR
  - Project counts (Sp/NF)
  - Project download counts (Sd/NF)
Regression Analysis – Definitions

• Linear regression obtains a linear relation between the dependent and the independent variables.

• The strength of the relation is represented by statistical quantities, such as:
  - Coefficient of determination, $R^2$
  - Sample correlation coefficient, $R$

• The significance of the relation is expressed by the level of significance $\alpha$, which is a measure of the chance that such a relation is NOT significant.
### Regression Analysis – Data

<table>
<thead>
<tr>
<th>Department no.</th>
<th>Undergraduate student repository activity (independent variable)</th>
<th>Faculty research impact indicator (dependent variable)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Project count (Sp/NF)</td>
<td>Download count (Sd/NF)</td>
</tr>
<tr>
<td>1</td>
<td>4.46</td>
<td>3,863.0</td>
</tr>
<tr>
<td>2</td>
<td>2.67</td>
<td>5,038.7</td>
</tr>
<tr>
<td>3</td>
<td>0.20</td>
<td>155.2</td>
</tr>
<tr>
<td>4</td>
<td>0.22</td>
<td>272.8</td>
</tr>
<tr>
<td>5</td>
<td>1.10</td>
<td>522.6</td>
</tr>
<tr>
<td>6</td>
<td>1.36</td>
<td>1,286.6</td>
</tr>
</tbody>
</table>

Most data are normalized by the department size represented by the number of faculty, $NF$. 

---

**CNI Meeting, December 2017**
Regression Analysis – Test for Normality

Quantile-quantile plots

Sample quantiles vs. Theoretical quantiles for:
- JNCI$_{av}$
- Sp/NF
- Sd/NF
- log(JNCI$_{av}$)
- log(Sp/NF)
- log(Sd/NF)
Regression Analysis – Results

Effect of Student Project Counts on JNCl\textsubscript{av}

- Observed data
- Regression line

All results are significant at $\alpha = 5\%$ level

$R^2 = 88\%$
$R = 0.94$
Regression Analysis – Results

Effect of Student Project Downloads on JNCI_{av}

- Observed data
- Regression line

All results are significant at $\alpha = 5\%$ level

$R^2 = 74\%$

$R = 0.86$
Regression Analysis – Results

Effect of Student Project Count on $\sum JNCl$

- Observed data
- Regression line

$R^2 = 74\%$

$R = 0.86$

All results are significant at $\alpha = 5\%$ level
Regression Analysis – Results

All results are significant at \( \alpha = 15\% \) level.
## Regression analysis results

<table>
<thead>
<tr>
<th>Statistics from regression analysis</th>
<th>Regression between log(JNCI_{av}) and log(Sp/NF)</th>
<th>Regression between log(JNCI_{av}) and log(Sd/NF)</th>
<th>Critical values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength of linear relationship</td>
<td>p value</td>
<td>p value</td>
<td>p_{\text{max}} = 0.05</td>
</tr>
<tr>
<td></td>
<td>p value</td>
<td>p value</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.006 &lt; 0.05</td>
<td>0.029 &lt; 0.05</td>
<td></td>
</tr>
<tr>
<td>Sample correlation coefficient, R</td>
<td>0.937 &gt; 0.812</td>
<td>0.859 &gt; 0.812</td>
<td>R \downarrow_{5%,6%} \uparrow_{\text{min}} = 0.812</td>
</tr>
<tr>
<td>Confidence in regression parameters</td>
<td>t statistics for a_0</td>
<td>t statistics for a_1</td>
<td>t_{4,2.5%} = 2.776</td>
</tr>
<tr>
<td></td>
<td>3.215 &gt; 2.776</td>
<td>5.346 &gt; 2.776</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.215 &gt; 2.776</td>
<td>5.346 &gt; 2.776</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Faculty repository activity

• Could faculty repository activity be the determining factor for their publications research impact?

• Could the strong correlation observed be the result of student repository activity being strongly correlated to faculty repository activity?

• A series of linear and bilinear regression analyses, involving the faculty repository activity along with student repository activity have been conducted to answer these questions.
Effect of Faculty Activity in IR

Bilinear regression analysis

- Independent variables (normalized by NF):
  - Faculty paper counts in IR (Fp)
  - Student project counts in IR (Sp)
- Dependent variable: $\text{JNCI}_{av}$

Data used in bilinear regression analysis

<table>
<thead>
<tr>
<th>Department no.</th>
<th>Activity in IR (independent variables)</th>
<th>$\text{JNCI}_{av}$ (dependent variable)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Student project counts (Sp/NF)</td>
<td>Faculty paper counts (Fp/NF)</td>
</tr>
<tr>
<td>1</td>
<td>4.46</td>
<td>14.91</td>
</tr>
<tr>
<td>2</td>
<td>2.67</td>
<td>3.05</td>
</tr>
<tr>
<td>3</td>
<td>0.20</td>
<td>1.57</td>
</tr>
<tr>
<td>4</td>
<td>0.22</td>
<td>3.56</td>
</tr>
<tr>
<td>5</td>
<td>1.10</td>
<td>16.17</td>
</tr>
<tr>
<td>6</td>
<td>1.36</td>
<td>11.55</td>
</tr>
</tbody>
</table>
Effect of Faculty Activity in IR

**Check for severe multicollinearity**

- Correlation coefficient between Fp and Sp: 
  \[ R_{Fp-Sp} = 0.29 \] is not significant \( \rightarrow \) no severe multicollinearity

**Effect on JNCI_{av}**

- Look at values of Adjusted \( R^2 \)

  - From linear regression analyses
    - \( \text{Adj.} R^2_{Sp-JNCI_{av}} = 85\% \)
    - \( \text{Adj.} R^2_{Fp-JNCI_{av}} = 44\% \)
  - From bilinear regression analysis
    - \( \text{Adj.} R^2 = 86\% \)
      - Effect of adding Sp=86%-44%=42\%
      - Effect of adding Fp=86%-85%=1\%
## Sample correlation coefficients

<table>
<thead>
<tr>
<th>Data pairs</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(Sp/NF) and log(JNCI$_{av}$)</td>
<td>0.937</td>
</tr>
<tr>
<td>log(Fp/NF) and log(JNCI$_{av}$)</td>
<td>0.741</td>
</tr>
<tr>
<td>log(Sp/NF) and log(Fp/NF)</td>
<td>0.632</td>
</tr>
<tr>
<td>log(Sd/NF) and log(JNCI$_{av}$)</td>
<td>0.859</td>
</tr>
<tr>
<td>log(Fd/NF) and log(JNCI$_{av}$)</td>
<td>0.625</td>
</tr>
<tr>
<td>log(Sd/NF) and log(Fd/NF)</td>
<td>0.290</td>
</tr>
</tbody>
</table>
## Adjusted R-square

<table>
<thead>
<tr>
<th>Regression analysis</th>
<th>AdjR²</th>
<th>Effect of adding factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Between $\log(\text{Sp/NF})$ and $\log(\text{JNCI}_{av})$</td>
<td>85%</td>
<td>$86% - 44% = 42%$</td>
</tr>
<tr>
<td>2. Between $\log(\text{Fp/NF})$ and $\log(\text{JNCI}_{av})$</td>
<td>44%</td>
<td>$86% - 85% = 1%$</td>
</tr>
<tr>
<td>3. Between $\log(\text{Sp/NF})$ &amp; $\log(\text{Fp/NF})$, as independent variables, and $\log(\text{JNCI}_{av})$, as dependent variable</td>
<td>86%</td>
<td></td>
</tr>
<tr>
<td>4. Between $\log(\text{Sd/NF})$ and $\log(\text{JNCI}_{av})$</td>
<td>67%</td>
<td>$82% - 24% = 58%$</td>
</tr>
<tr>
<td>5. Between $\log(\text{Fd/NF})$ and $\log(\text{JNCI}_{av})$</td>
<td>24%</td>
<td>$82% - 67% = 15%$</td>
</tr>
<tr>
<td>6. Between $\log(\text{Sd/NF})$ &amp; $\log(\text{Fd/NF})$, as independent variables, and $\log(\text{JNCI}_{av})$, as dependent variable</td>
<td>82%</td>
<td></td>
</tr>
</tbody>
</table>
Faculty repository activity

**Results**

- Correlation between faculty repository activity and research impact resulted weaker than the correlation between student repository activity and research impact.

- Student repository activity contribute more significantly to the bilinear regression model for predicting research impact than IR faculty activity.
Impact of faculty research can be correlated with the student research activity in DC@CalPoly.

Undergraduate student senior projects follow (and sometimes anticipate) the topics of faculty research.

Causality between student research quality and faculty research quality.

Evidence Based Library and Information Practice, vol. 12 (4), December 2017
## Next steps

<table>
<thead>
<tr>
<th>Expand to other disciplines and bigger sample</th>
<th>Effect of master theses IR exposure</th>
<th>Correlation between individual faculty research impact and student advisees’ IR activity</th>
</tr>
</thead>
</table>

Correlation between undergraduate student project deposits and faculty research impact
Reactions? Questions? Suggestions?