Validation Sample Studies and Further Analyses for EAST

prepared by: Grant Ritter, The Heller School, Brandeis University
date: October 14, 2016
Primary Goal of Validation Sample Study 1

- Develop a methodological approach to the study
- Determine rates of missing monographs at participating EAST libraries
- Identify factors which predict higher risk for being missing
- Determine if there are necessary modifications to the current EAST retention program to guard against accidentally ‘losing’ last monograph copy
Secondary Goal of Validation Sample Study 1

- Provide statistical analyses concerning the condition of monographs
- Determine distribution in the conditions of monographs at each EAST library - poor, acceptable or excellent
- Identify factors which predict higher risk for being in poor condition
- Determine any necessary modifications of the current EAST retention program to guard against all retained copies being in poor condition
Data Collection Plan

- Draw a random sample of monographs from each participating EAST library.
- Have staff determine status of each monograph on each library’s sample list,
- Status should be recorded as present in stacks, checked-out, or LMBO (lost, missing, billed or otherwise unavailable)
- Searcher should also record the condition of each monograph found present in the stacks
Question of Accuracy

- Question: What accuracy in missing rate do we need, and how many monographs at each library should be sampled to achieve it?
- Desired accuracy has established as 1% by Monographs Working Group
- Sample size to assure 1% accuracy does not depend on the number of monographs in library
- Sample size to assure 1% accuracy does depend on the underlying rate of missing at the library
Calculation of Sample Size

- 1% accuracy for any baseline missing rate requires 16,000 monographs per library
- However, if missing rate at library is less than 10%, then only 6,000 monographs are required
- In the Mellon proposal it was estimated that 10% is likely to be upper limit for rates of missing and the Executive Committee approved 6,000 monographs as sample size
- Consequence for a missing rate above 10% is to realize a bit less accuracy (like 2% instead of 1%)
Data Collection

- Individual random samples of 6,000 monographs were selected for the participating EAST libraries.
- Libraries trained helpers to search for and determine the status of each sampled monograph (also the condition, if present in the stacks).
- Monograph status was recorded as present, checked-out, or LMBO.
- Searchers were equipped with bar code scanners to assure accuracy.
Initial Analyses – per Individual Library

- Calculation of percentage of monographs missing for each participating EAST library
- Calculation of percentage of monographs in poor, acceptable, or excellent condition for each EAST library
Initial Analyses – Individual Libraries

- Summary Statistics and Distribution on Missing Monographs at 40 participating EAST libraries (all estimates accurate to 1%)

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std</th>
<th>Min</th>
<th>5th pctl</th>
<th>25th pctl</th>
<th>50th pctl</th>
<th>75th pctl</th>
<th>95th pctl</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>3.0%</td>
<td>2.1%</td>
<td>.3%</td>
<td>.6%</td>
<td>1.3%</td>
<td>2.8%</td>
<td>4.2%</td>
<td>7.4%</td>
<td>9.7%</td>
</tr>
</tbody>
</table>
Initial Analyses – Individual Libraries

- Summary Statistics and Distribution on Condition of Monographs at 40 participating EAST libraries (all estimates accurate to 2%):

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std</th>
<th>Min</th>
<th>5th pctl</th>
<th>25th pctl</th>
<th>50th pctl</th>
<th>75th pctl</th>
<th>95th pctl</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>10.7%</td>
<td>9.8%</td>
<td>1.1%</td>
<td>1.3%</td>
<td>3.0%</td>
<td>8.2%</td>
<td>15.2%</td>
<td>29.3%</td>
<td>44.2%</td>
</tr>
<tr>
<td>Acceptable</td>
<td>55.7%</td>
<td>17.3%</td>
<td>34.8%</td>
<td>34.8%</td>
<td>45.5%</td>
<td>54.1%</td>
<td>65.9%</td>
<td>88.1%</td>
<td>91.1%</td>
</tr>
<tr>
<td>Excellent</td>
<td>36.6%</td>
<td>17.7%</td>
<td>6.8%</td>
<td>7.8%</td>
<td>20.7%</td>
<td>33.6%</td>
<td>43.3%</td>
<td>70.5%</td>
<td>78.8%</td>
</tr>
</tbody>
</table>
Follow-up Analyses – Factors Affecting Missing

- Is missing related to the age of the monograph?
- Is missing related to the frequency of use?
- Is missing related to the monograph’s level of duplication across EAST libraries?
- Are certain subject areas more prone to missing?
- Are any of these factors important enough to modify the current EAST retention plan?
Logistic Regression—Significant Factors

- Models being missing in terms of factors which might impact its likelihood: age, frequency of circulation, level of duplication across EAST, subject area

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of monograph</td>
<td>.07%</td>
<td>4.73</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Frequency of Circulation</td>
<td>1.6%</td>
<td>8.64</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Number of duplicates within EAST</td>
<td>-.17%</td>
<td>-7.05</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Call number for US Law (‘KF’)</td>
<td>1.7%</td>
<td>4.41</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Call number for Mathematics (‘QA’)</td>
<td>.56%</td>
<td>2.06</td>
<td>.0396</td>
</tr>
</tbody>
</table>

1. estimate of effect for 10 year increase in age of item
2. estimate of effect for increase of 20 in frequency of use
3. estimate of effect for increase in 5 in number of duplicate items in EAST libraries
Follow-up Analyses – Factors Affecting Condition of Monograph

- Is condition related to the age of the monograph?
- Is condition related to the frequency of use?
- Is condition related to the monograph’s level of duplication across EAST libraries?
- Are certain subject areas more prone to being in poor condition?
- Are any of these factors important enough to modify the current EAST retention plan?
Logistic Regression– Significant Factors

- Models being in poor condition in terms of age, frequency of circulation, level of duplication across EAST, subject area

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of monograph(^1)</td>
<td>3.0%</td>
<td>124.1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Frequency of Circulation(^2)</td>
<td>5.1%</td>
<td>14.61</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Number of duplicates within EAST(^3)</td>
<td>-.4%</td>
<td>-9.32</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Call number for African History (‘DT’)</td>
<td>1.9%</td>
<td>2.72</td>
<td>.007</td>
</tr>
<tr>
<td>Call number for Paintings (‘ND’)</td>
<td>3.2%</td>
<td>5.92</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

\(^1\) estimate of effect for 10 year increase in age of item  
\(^2\) estimate of effect for increase of 20 in frequency of use  
\(^3\) estimate of effect for increase in 5 in number of duplicate items in EAST libraries
Conclusions – Likelihood of Being Missing

- Missing rates at most participating libraries were very low. Only two libraries had rates greater than 7.4%, and no library had a rate greater than 10%.

- Factors significantly affecting likelihood of being missing were small.

- No factor appeared important enough to suggest a modification to the current EAST retention plan.
Conclusions – Likelihood of Being in Poor Condition

- Poor condition rates at some participating libraries were large enough to merit attention. A quarter of the libraries had rates greater than 15% and two libraries had rates greater than 29%.

- Certain factors affecting likelihood of being in poor condition are large enough to recommend changes to retention plans.

- Retention plan might keep extra copies of older monographs, monographs frequently circulated, and monographs in subject area of painting.
Next Steps

- A statistical analysis will simulate the current retention program to check whether the current validation study data predict any problems with the plan.

- Retention model predicted retention of approximately 36% of monographs. Of the 240,000 monographs in the sample, 92,500 are identified for retention.

- Findings with respect to missing or poor condition monographs among these 92,500 may inform us of possible issues with the retention model which would influence subsequent modeling, particularly for a new cohort.