Validation sampling as part of a shared print program: EAST

Susan Stearns, EAST Project Director
ALA Annual PAN Meeting, Friday, June 23, 2017
www.eastlibraries.org
sstearns@blc.org
The EAST mission

Securing the scholarly print record in support of teaching, learning and research
- 59 members from Maine to Florida
- 51 Retention Partners
- 8 Supporting Partners
- Diversity in collection size, research vs. teaching focus and collection philosophy
Validation is infrequent

• Survey by Terri Koch and Andrew Welch of Drake University in 2016 of 14 shared print programs
  • Mix of serials and monographs
  • 12 either doing no validation or still considering
  • Reasons for NOT validating: Time and Cost
  • Reasons for validating: Ensure access, reassure faculty

• EAST: our funder cared a great deal and provided monies to cover a sampling approach
Original Goals for the Validation Sample Study

- Develop study methodology
- Determine missing rates
- Determine factors impacting “missingness”
Further Goals Added

- Cursory condition evaluation
- Determine factors correlated with poor condition
- Should retention model or allocations be modified?
Our Methodology

1. Draw sample
2. ILS check
3. Shelf check
4. Condition check
5. Upload data
EAST Data Collection Tool

https://github.com/samato88/EastValidationTool
Summary Statistics and Distribution on Missing Monographs
(all estimates accurate to 1%)

<table>
<thead>
<tr>
<th>Rate</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>

= 97% average availability metric
Summary Statistics and Distribution on Condition of Monographs
(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>

= almost 90% in good or excellent condition
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 25%.

- 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.
Deeper Dive into the Data

• Statistical analysis to model likelihood of item being missing or in poor condition
• Identified 7,800 items with statistically higher likelihood of being missing
• Identified 72,700 items with statistically higher likelihood of being in poor condition – over half are pre-1900 imprints
• Total represents less than .01% of collective collection
• Proposed a mitigation strategy to retain an additional copy
Current Status

- Libraries taking on additional retention commitments based on the “deeper dive”
- Push retention commitments to Cohort 2 when possible
- Replicate the sample study with 12 Cohort 2 libraries – including 2 with collection sizes of more than 3 million
- Refinement of documentation and training for Cohort 2 libraries
- Determine if results should impact retention model or allocation
  - Should we retain additional copies of older materials?
  - Should we avoid allocation to libraries with low availability?
Thank you.

sstearns@blc.org