**Validation Data Analyses: Missing Status and Assessed Condition of Monographs at EAST Libraries**
**Prepared by Grant A. Ritter, Brandeis University**
**Date: June 9, 2018**

**Overview**: This study continues the investigation of the rates of missing of monographs at EAST libraries, together with the condition of monographs in the stacks, which therefore can be assessed. Following the original analyses of monographs at 40 college and university libraries in the EAST consortium (now referred to as 'cohort 1' of EAST), an additional group of 12 libraries (now referred to as 'cohort 2') joined EAST and consented to have their monographs examined as well. This report documents the results of this new undertaking. Like the original EAST study, the analyses used randomly selected samples of 6,000 monographs per cohort 2 library, the same as in cohort 1. The study calculates statistics on rates of missing and condition of cohort 2 monographs alone, and also across the combined group of all 52 EAST libraries. Following this, regression models are performed to determine the properties of monographs which increase their likelihood of being missing or in poor condition. These models are first performed using the cohort 2 sample of 72,000 monographs alone, but then are also run using the combined sample of cohort 1 and cohort 2 monographs together (n=312,000).

**Study design and Power:** At each participating cohort 2 library in the current study, sampled monographs first had their checked-out status determined in the library's circulation database. Monographs reported as present in the stacks were searched for, and if found, received an assessment of condition. All monographs were then classified as present, currently checked out, or LMBO (i.e., 'loss, missing, billed, or otherwise unavailable'), and assessed monographs received a condition value of excellent, acceptable, or poor. The number of monographs per library was chosen to provide an accurate estimate of the rate of missing. Provided the missing rate was 10% or less, a sample size of 6,000 would provide a 99% likelihood of an estimate within 1% (e.g., if the observed missing rate were 4.0% within the sample, we would be 99% confident that the actual rate of missing at the library was between 3.0% and 5.0%).

**Methods:** The descriptive analysis component of the current study calculated rates of missing at each library and frequency distributions on their assessed conditions. Statistics of the individual libraries were then further analyzed to provide overall mean and variation. The multivariate analysis component of the study then combined the 6,000 monographs per library to form two samples: a cohort 2 only sample of 72,000 volumes and a sample combining cohort 1 and cohort 2 resulting in 312,000 volumes. Logistic models were conducted on each sample to identify the factors which predicted being missing, and then being in poor condition. The factors used in these logistic models include a scale[1] for the age of the item (increasing by one unit for each ten years since publication date), a scale for the frequency of circulation (increasing by one unit for each 20 recorded uses), a scale for the number of copies of the title across all US libraries, (increasing by one unit for each five copies), and subject area (as represented by the two-character call number).

---

[1] The models would be just as valid if, instead of scales, the original variables in the data, like the age of the item in years or the number of recorded uses, were used as the independent variables. Scales generate differences in the expected likelihoods of the outcomes which are easier to interpret.

**Description Results**: Table 1 provides percentile values on the rates of missing among library cohorts and then for the 52 EAST libraries as a single group (note: Appendix 1 provides the individual rates of missing among the 52 library samples). Table 1 results, which are calculated based on equal weights for all participating libraries, show mean estimated rates of missing equal to 3.0% for cohort 1 libraries, 2.18% for cohort 2 libraries, and 2.79% for the combined group. All estimates are accurate to within 1.0% with 99% likelihood. The table also shows that the distribution of estimated rates of missing has a 90th percentile value of 7.4% for cohort 1 libraries, 4.4% for cohort 2 libraries, and 5.22% across all 52 EAST libraries (indicating that only five libraries had missing rates of 5.22% or more). Similarly, the 75th percentile rate of missing was 4.2% for cohort 1 libraries, 2.9% for cohort 2 libraries, and 3.6% for the combined group (indicating that three quarters, or 39 libraries, had a rate of missing of 3.6% or less). Notably, all participating libraries had estimated rates of missing under 10.0%, indicating strong likelihood that all estimated rates of missing were accurate to within 1.0%.

**Table 1: Rates of Missing Monographs – by Cohort and Overall Among all 52 EAST Libraries**

|          | mean  | std[1] | 5th pctl | 10th pctl | 25th pctl | 50th pctl | 75th pctl | 90th pctl | 95th pctl |
|----------|-------|-------|---------|----------|----------|----------|----------|----------|----------|
| Cohort 1 | 3.0%  | 2.1%  | .3%     | .6%      | 1.3%     | 2.8%     | 4.2%     | 7.4%     | 9.7%     |
| Cohort 2 | 2.18% | 1.55% | .25%    | .77%     | .86%     | 1.85%    | 2.89%    | 4.40%    | 5.42%    |
| Combined | 2.79% | 2.03% | .3%     | .63%     | 1.18%    | 2.44%    | 3.60%    | 5.22%    | 6.47%    |

[1] The standard deviation (std) gives some idea of the spread on the values. Roughly 80% of all library rates should be with one std of the mean, and roughly 95% of all rates should be within two std of the mean.

Table 2 provides summary results on the distributions for the conditions of the monographs at the two library cohorts and then as a combined group (note: Appendix 2 provides the individual condition proportions for all 52 EAST libraries). Table 2, again based on equal weights, indicates that the mean poor condition proportion was 10.4% among cohort 1 volumes, 7.1% among cohort 2 volumes, and 9.94% for the combined group. Similarly, the mean acceptable condition proportion was 55.7% among cohort 1 volumes, 46.9% among cohort 2 volumes, and 54.1% for the combined group, while the mean excellent condition proportion was 36.6% among cohort 1 volumes, 42.2% among cohort 2 volumes, and 36.0% for the combined group. Estimates of poor condition monographs at all libraries were accurate to within 1% with 99% likelihood. Table 2 also provides the distributions for various percentile values, including that the 10th and 90th percentiles for poor condition titles were 1.09% and 21.5% respectively (indicating that all but 11 libraries had between 1.09% and 21.5% of their titles in poor condition). It can be noted that the variation in poor condition monographs among the 52 libraries is much greater than the variation in missing. This is due in part because the mean rate of poor condition is higher than the mean rate of missing, but it is likely also due to the fact that judging a monograph to be in poor condition is subjective and there could be significant variation among reviewers in what constituted poor condition.

**Table 2: Distribution of Condition of Monographs - by Cohort and Overall Among all 52 EAST Libraries**

| | mean | std | 5th pctl | 10th pctl | 25th pctl | 50th pctl | 75th pctl | 90th pctl | 95th pctl |
|---|---|---|---|---|---|---|---|---|---|
| Cohort 1 | | | | | | | | | |
| Poor | 10.7% | 9.8% | 1.31% | 1.54% | 3.04% | 8.19% | 15.2% | 22.8% | 29.3% |
| Acceptable | 55.7% | 17.3% | 28.2% | 34.8% | 45.5% | 54.1% | 65.9% | 84.1% | 91.0% |
| Excellent | 33.6% | 17.7% | 7.84% | 12.7% | 20.7% | 33.6% | 43.3% | 57.1% | 70.5% |
| | | | | | | | | | |
| Cohort 2 | | | | | | | | | |
| Poor | 7.1% | 5.9% | .68% | .83% | 2.20% | 5.55% | 11.2% | 14.8% | 18.8% |
| Acceptable | 46.9% | 18.4% | 3.2% | 38.2% | 42.5% | 45.6% | 53.7% | 61.2% | 84.1% |
| Excellent | 42.2% | 20.0% | 7.9% | 30.8% | 33.2% | 41.0% | 47.0% | 50.9% | 94.6% |
| | | | | | | | | | |
| Overall | | | | | | | | | |
| Poor | 9.94% | 9.09% | 1.09% | 1.54% | 2.89% | 7.96% | 14.5% | 21.5% | 26.6% |
| Acceptable | 54.1% | 17.9% | 27.4% | 35.2% | 45.4% | 52.2% | 60.4% | 82.4% | 89.9% |
| Excellent | 36.0% | 18.6% | 7.88% | 13.7% | 23.0% | 34.8% | 45.8% | 52.7% | 71.0% |

**Factors Affecting Likelihood of Being Missing**: The study also attempted to determine factors which predicted the likelihood of monographs being missing. Table 3 provides results of such multivariate logistic regressions using variables for age, frequency of circulation, number of titles of the monograph in the US, and subject area (as represented by a two-character call number). Individual indicators for each EAST library were also included in the model specification. These models were conducted using the cohort 2 sample only (n=72,000) and then using the combined sample (n=312,000 volumes). Results of the two regression models were very similar and showed that after controlling for library, the only consistently significant predictors for an item being missing were the age of the monograph and having its subject matter classified as Religion ('BL') or US Law ('KF'). In particular, US Law monographs had an odds ratio of approximately 4.5 for being missing (for example, a 2% likelihood would become a 9% likelihood) and Philosophy and Religion monographs demonstrated an odds ratio of approximately 1.8 for being missing. In addition, every 10 year increase in the age of an item associates with a 4% to 5% increase in its likelihood of being missing. Since some monographs are greater than 50 years old, this factor could predict a much higher likelihood (e.g., 20%-25% higher) compared with fairly new items. However, both models were consistent in noting the strongest characteristic increasing the likelihood of a monograph being missing was the library itself. As noted in the original EAST study, volumes at Bridgewater State, UMass Dartmouth, and UMass Lowell were much more likely to be missing compared with other cohort 1 libraries. Among Cohort 2 libraries, those at New York University and University of Pittsburgh shared this problem.

**Table 3: Factors Predicting Likelihood of Being Missing**

| Variable | Cohort 2 Only | | | Combined Sample (1 and 2) | | |
|---|---|---|---|---|---|---|
| | Odds Ratio | t-value | p-value | Odds Ratio | t-value | p-value |
| Age of monograph (per year)[1] | 1.05 | 6.12 | <.0001 | 1.04 | 6.07 | <.0001 |
| Call number for Religion ('BL') | 1.82 | 2.46 | .0137 | 1.83 | 2.35 | .0185 |
| Call number for US Law ('KF') | 4.43 | 9.74 | <.0001 | 4.65 | 9.96 | <.0001 |
| Bridgewater State | NA | - | - | 1.58 | 18.2 | <.0001 |
| UMass Dartmouth | NA | - | - | 1.84 | 22.0 | <.0001 |
| UMass Lowell | NA | - | - | 2.82 | 35.4 | <.0001 |
| New York University | 1.10 | 15.8 | <.0001 | 1.31 | 12.7 | <.0001 |
| University of Pittsburgh | 1.36 | 20.3 | <.0001 | 1.53 | 16.2 | <.0001 |

[1] estimate of effect per 10 year increase in age of item

**Factors Affecting Likelihood of Being in Poor Condition**: Similar logistic models were used to determine factors which predict monographs in poor condition. Table 4 provides the results of these models, first based on Cohort 2 only and then using the combined sample. These logistic results are somewhat different from missing, in that they have a much larger number of significant predictors.  Based on the Cohort 2 sample, each ten year increase in the age of the item, each increment of twenty additional checkouts of a monograph, and each increment of 5 in the number of US holdings leads to an odds ratio of being in poor condition of 1.036, 1.05, and 1.038 respectively. In addition, a larger set of call numbers are indicative of monographs being in poor condition. For example, monographs in the areas of Psychology (OR=1.48), Asian History (OR=1.38), Economic History (OR=1.53), Family and Marriage (OR=1.70), Welfare and Criminology (OR=2.31), Theory and Practice of Education (OR=1.63), Painting (1.61), and French and Spanish Literature (1.28) all are more likely to have monographs in poor condition. Almost all of these factors were also significantly in the logistic models using the combined sample of Cohorts 1 and 2 and had similar odds ratios. Only two factors, number of US holdings and being concerned with French or Spanish literature (I.e., call number 'PQ'), did not retain their significance with the combined sample. At the same time, only one factor, being concerned with English literature (i.e., call number 'PR') gained significance in the larger combined sample.

Just as with the outcome of being missing, library indicators were among the strongest predictors for monographs being in poor condition. In this regard, cohort 2 libraries at Bucknell (OR 2.11), Florida State (OR 1.46), and Gettysburg (OR 3.28) joined cohort 1 libraries at Amherst, Bard, Bryn Mawr, Hampshire, Mount Holyoke, Skidmore, Trinity, Union, and Wellesley in having the highest estimated odds ratios. As noted in the first EAST study, one caveat to all of the analyses of monograph conditions is that the assessments at different libraries were done by different reviewers. Although reviewers were given training, it is likely they employed somewhat different standards for determining monograph condition. This is particularly likely for libraries such as at Amherst and Skidmore colleges, for whom a much larger proportion of monographs were judged to be in poor condition.

**Table 4: Estimated Effects Predicting Monographs in Poor Condition**

| Variable | Cohort 2 Only | | | Combined Sample | | |
|---|---|---|---|---|---|---|
| | Odds Ratio | t-value | p-value | Odds Ratio | t-value | p-value |
| Age of monograph[1] | 1.036 | 65.8 | <.0001 | 1.035 | 65.8 | <.0001 |
| Frequency of Circulation[2] | 1.05 | 23.1 | <.0001 | 1.038 | 33.0 | <.0001 |
| Number of US holdings[3] | 1.038 | 4.64 | <.0001 | NS | - | - |
| Call number for Psychology ('BF') | 1.48 | 2.61 | .009 | 1.60 | 2.98 | .003 |
| Call number for Asian History ('DS') | 1.38 | 2.50 | .013 | 1.43 | 2.48 | .013 |
| Call number for Economic History ('HC') | 1.53 | 2.78 | .005 | 1.53 | 2.51 | .012 |
| Call number for Family and Marriage ('HQ') | 1.70 | 4.16 | <.0001 | 1.85 | 4.65 | <.0001 |
| Call number for Welfare and Criminology ('HV') | 2.31 | 6.90 | <.0001 | 2.48 | 7.29 | <.0001 |
| Call number for Theory and Practice of Education ('LB') | 1.63 | 3.05 | .002 | 1.65 | 2.90 | .004 |
| Call number for Paintings ('ND') | 1.61 | 3.07 | .002 | 1.68 | 3.12 | .002 |
| Call number for French and Spanish Literature ('PQ') | 1.28 | 2.14 | .032 | NS | - | - |
| Call number for English Literature ('PS') | NS | - | - | 1.36 | 2.81 | .005 |
| | | | | | | |
| Bucknell University | 2.10 | 27.5 | <.0001 | 2.11 | 21.0 | <.0001 |
| Florida State University | 1.38 | 40.2 | <.0001 | 1.46 | 11.0 | <.0001 |
| Gettysburg College | 3.06 | 16.7 | <.0001 | 3.28 | 36.1 | <.0001 |
| | | | | | | |
| Amherst College | NS | - | - | 11.04 | 88.5 | <.0001 |
| Bard College | NS | - | - | 3.68 | 42.3 | <.0001 |
| Bryn Mawr University | NS | - | - | 3.37 | 39.8 | <.0001 |
| Hampshire College | NS | - | - | 3.18 | 37.0 | <.0001 |
| Mount Holyoke College | NS | - | - | 3.39 | 40.2 | <.0001 |
| Skidmore College | NS | - | - | 6.14 | 65.4 | <.0001 |
| Trinity College | NS | - | - | 2.66 | 30.7 | <.0001 |
| Union College | NS | - | - | 4.30 | 51.0 | <.0001 |
| Wellesley College | NS | - | - | 2.51 | 28.7 | <.0001 |

[1] estimate of effect per 10 year increase in age of item
[2] estimate of effect for increase in frequency of use
[3] estimate of effect for increase in US holdings by 100

**Conclusions:** The results of this new study, which adds a new cohort of libraries to the EAST consortium, confirm many of the findings and patterns found in the first study. For the most part, the likelihood of a monograph being missing is low. There is, however, variation in the likelihood of being missing based on the monograph's subject matter, and even more so, on where the monograph is housed. Factors such as the age of the monograph and how often it is checked out also influence the likelihood of being missing, but only in a minor way.

With regard to the condition of monographs at the EAST libraries, again there is significant variation based on age and frequency of use, but again the subject area of the monograph and where it is housed prove to be even more significant. In particular, three additional cohort 2 libraries, Bucknell, Florida State, and Gettysburg, had enough monographs in poor condition (11% or more), that it might be worthwhile to adjust for their higher likelihood of having monographs in poor condition in developing any future EAST retention plan.

**Appendix 1a: Rates of Missing Monographs Among Cohort 2 EAST Libraries**

| Library | N Obs | Rate of Missing | Std Error |
|---|---|---|---|
| Bucknell University | 6000 | 1.65% | 0.16% |
| Davidson College | 6000 | 1.85% | 0.17% |
| DeSales University | 6000 | 0.77% | 0.11% |
| Florida State University | 6000 | 3.33% | 0.23% |
| Furman University | 6000 | 2.45% | 0.20% |
| Gettysburg College | 6000 | 0.77% | 0.11% |
| Hofstra University | 6000 | 1.85% | 0.17% |
| New York University | 6000 | 4.40% | 0.27% |
| Syracuse University | 6000 | 2.43% | 0.20% |
| Union College (cohort 2) | 6000 | 0.25% | 0.06% |
| University of Pittsburgh | 6000 | 5.42% | 0.29% |
| University of the South | 6000 | 0.95% | 0.13% |

**Appendix 1b: Rates of Missing Monographs Among Cohort 1 EAST Libraries**

| Library | N Obs | Rate of Missing | Std Error |
|---|---|---|---|
| Adelphi University | 6000 | 3.2% | 0.2% |
| Amherst College | 6000 | 4.7% | 0.3% |
| Bard College | 6000 | 8.7% | 0.4% |
| Boston College | 6000 | 5.1% | 0.3% |
| Boston University | 6000 | 3.0% | 0.2% |
| Brandeis University | 6000 | 3.3% | 0.2% |
| Bridgewater State | 6000 | 5.7% | 0.3% |
| Brynmawr University | 6000 | 2.7% | 0.2% |
| Colby College | 6000 | 1.0% | 0.1% |
| Connecticut College | 6000 | 2.6% | 0.2% |
| Elms College | 6000 | 4.7% | 0.3% |
| Fairfield University | 6000 | 3.1% | 0.2% |
| Hamilton College | 6000 | 0.6% | 0.1% |
| Hampshire College | 6000 | 2.1% | 0.2% |
| Haverford College | 6000 | 1.4% | 0.1% |
| Lafayette College | 6000 | 0.6% | 0.1% |
| Loyola University | 6000 | 3.2% | 0.2% |
| Middlebury College | 6000 | 2.3% | 0.2% |
| Mount Holyoke College | 6000 | 1.6% | 0.2% |
| Phillips Exeter Academy | 6000 | 1.7% | 0.2% |
| Siena College | 6000 | 0.3% | 0.1% |
| Skidmore College | 6000 | 1.2% | 0.1% |

| Smith College | 6000 | | 4.9% | 0.3% |
|---|---|---|---|---|
| St Anselm College | 6000 | | 1.1% | 0.1% |
| Swarthmore College | 6000 | | 1.0% | 0.1% |
| Trinity College | 6000 | | 2.4% | 0.2% |
| Tufts University | 6000 | | 3.1% | 0.2% |
| University Connecticut | 6000 | | 5.2% | 0.3% |
| UMass Amherst | 6000 | | 1.8% | 0.2% |
| UMass Boston | 6000 | | 3.2% | 0.2% |
| UMass Dartmouth | 6000 | | 6.5% | 0.3% |
| UMass Lowell | 6000 | | 9.0% | 0.4% |
| University of New Hampshire | 6000 | | 1.1% | 0.1% |
| Union College | 6000 | | 0.6% | 0.1% |
| University of Rochester | 6000 | | 3.1% | 0.2% |
| Vassar College | 6000 | | 4.7% | 0.3% |
| Wellesley College | 6000 | | 1.6% | 0.2% |
| Wesleyan University | 6000 | | 2.9% | 0.2% |
| Williams College | 6000 | | 0.5% | 0.1% |

**Appendix 2a: Distributions in Condition of Monographs Among Cohort 2 EAST Libraries**

| Library | N Obs[1] | % Poor | % Acceptable | % Excellent |
|---|---|---|---|---|
| Bucknell University | 5901 | 13% | 52% | 35% |
| Davidson College | 5889 | 0.9% | 3.3% | 95% |
| DeSales University | 5954 | 0.7% | 49% | 50% |
| Florida State University | 5800 | 11% | 47% | 43% |

| | | | | |
|---|---|---|---|---|
| Furman University | 5853 | 2.9% | 45% | 52% |
| Gettysburg College | 5954 | 19% | 39% | 42% |
| Hofstra University | 5889 | 7.9% | 61% | 31% |
| New York University | 5736 | 5.3% | 45% | 50% |
| Syracuse University | 5854 | 1.7% | 90% | 8.4% |
| Union College (cohort 2) | 5985 | 15% | 47% | 38% |
| University of Pittsburgh | 5675 | 4% | 49% | 47% |
| University of the South | 5943 | 7% | 59% | 34% |

[1] Number of items is less than 6,000 because books checked out or LMBO could not be examined.

**Appendix 2b: Distributions in Condition of Monographs Among Cohort 1 EAST Libraries**

| Library | N Obs[1] | % Poor | % Acceptable | % Excellent |
|---|---|---|---|---|
| Adelphi University | 5755 | 1.4% | 19.8% | 78.8% |
| Amherst College | 5511 | 47.2% | 38.7% | 14.2% |
| Bard College | 5210 | 24.2% | 39.5% | 36.3% |
| Boston College | 5129 | 10.5% | 56.7% | 32.8% |
| Boston University | 5662 | 12.2% | 59.5% | 28.3% |
| Brandeis University | 5548 | 3.2% | 66.9% | 29.9% |
| Bridgewater State | 5555 | 2.1% | 57.7% | 40.2% |
| Brynmawr University | 5616 | 21.5% | 41.6% | 36.9% |
| Colby College | 5703 | 1.1% | 29.0% | 69.9% |
| Connecticut College | 5720 | 1.5% | 27.4% | 71.0% |
| Elms College | 5697 | 14.1% | 64.9% | 21.0% |
| Fairfield University | 5780 | 7.9% | 48.5% | 43.6% |
| Hamilton College | 5716 | 8.2% | 56.9% | 34.9% |
| Hampshire College | 5639 | 2 % | 56.4% | 23.6% |
| Haverford College | 5654 | 15.4% | 50.3% | 34.3% |
| Lafayette College | 5786 | 3.3% | 35.2% | 61.6% |
| Loyola University | 5744 | 1.9% | 45.5% | 52.6% |
| Middlebury College | 5532 | 2.0% | 77.6% | 20.3% |
| Mount Holyoke College | 5663 | 21.5% | 50.5% | 28.0% |
| Phillips Exeter Academy | 5754 | 9.2% | 58.0% | 32.8% |
| Siena College | 5894 | 8.0% | 53.8% | 38.2% |
| Skidmore College | 5872 | 4.4% | 45.8% | 49.9% |
| Smith College | 5830 | 32.0% | 45.6% | 22.4% |
| St Anselm College | 5558 | 8.7% | 39.5% | 51.7% |
| Swarthmore College | 5631 | 8.7% | 54.4% | 36.9% |
| Trinity College | 5713 | 17.4% | 56.9% | 25.7% |
| Tufts University | 5694 | 8.2% | 51.5% | 40.3% |
| University Connecticut | 5834 | 26.6% | 49.8% | 23.5% |
| UMass Amherst | 5567 | 2.7% | 85.7% | 11.6% |
| UMass Boston | 5747 | 8.9% | 56.8% | 34.3% |
| UMass Dartmouth | 5724 | 1.5% | 90.5% | 7.9% |
| UMass Lowell | 5357 | 12.9% | 71.9% | 15.2% |
| University of New Hampshire | 5305 | 3.8% | 82.5% | 13.7% |
| Union College | 5641 | 15.0% | 71.3% | 13.7% |
| University of Rochester | 5661 | 2.9% | 53.9% | 43.3% |
| Vassar College | 5599 | 1.2% | 91.1% | 7.8% |
| Wellesley College | 5674 | 16.8% | 34.3% | 48.8% |
| Wesleyan University | 5111 | 4.6% | 52.2% | 43.2% |
| Williams College | 5803 | 7.7% | 72.7% | 19.6% |