

**College Library Statistics: Under Review**  
**Teresa A. Fishel**  
**Macalester College**

Iowa Private Academic Libraries  
March 22, 2007  
Mt. Mercy College, Cedar Rapids, Iowa

I need to state for the record that I am not an authority on statistics. I have no formal training and my qualifications as far as speaking on this topic are limited to my personal experiences collecting and reporting statistics for the various surveys academic libraries are asked to complete. These surveys merely serve as a means of recording numbers which may or may not provide measurements of certain activities that take place in our libraries. This year I started looking more closely at the statistics I was asked to supply and starting to ask questions as to why we were collecting certain numbers and not others. Currently I'm working with a small group to try and "retool" the statistics that are compiled annually for the Oberlin group. We want to address some of the inconsistencies in the data collection that have arisen over time due to technological and other changes that we've all experienced in academic libraries. So those are my qualifications as far as talking to you regarding statistics.

The fact that I'm not a statistician, doesn't mean I don't have some opinions on this annual practice of collecting data by submitting numbers that are supposed to represent our activities. As my staff and others will confirm, I'm not short on opinions, and since the annual completion of statistics is an area that does require a significant amount of staff time, it is something to which I have given some thought. This is especially true as we recently completed a review of our goals for the year and I needed to apply some means of measurement to assess our success in completing them. I'm sure that a number of you have been involved in either compiling or submitting numbers for either your own annual report, or for a regional or national survey. In fact I thought it was appropriate that I would be speaking today on the deadline for the Academic Libraries Survey (ALS) formerly part of the IPEDs. As I begin, I'd like you all to reflect on how much time you spend compiling numbers that contribute to your own annual reports, or regional or national surveys. Perhaps you delegate that assignment, so consider how much staff time is spent gathering the numbers and double-checking to make sure they are accurate. You collect the data, and then

what do you do with the numbers? I think the more critical aspect is what you do with the numbers after you've compiled them. I think most of us would agree that we are providing statistics in order to obtain comparisons from other institutions that are similar to us. Perhaps you feel that by comparing yourselves with other similar institutions, you can learn something about current trends. However, the question is, what is it you really learn? And, more importantly, what do you do with what you've learned? Are you using the numbers for the purpose of preparing reports for your administration to justify or request additions for allocations or staff? Are you identifying areas where you think you could be making process improvements in your own library? Are you analyzing trends for strategic planning? Does the study of those trends include changes in staff assignments and work flows? Has anyone identified something they could stop doing based on statistics they kept? Has anyone identified something they know they can't stop doing based on statistics they keep? Upon reflection, you may find the numbers aren't providing you with the data that you need to make decisions--which leads me to the subject of our conversation today. In an article that appeared in *Library Trends* in 2004, it was pointed out that this question was raised over twenty years ago:

“Do librarians collect the appropriate statistics? Are the statistics collected either accurate or comparable among similar libraries? Do we ask valid questions of the data? And above all, do we know how to manipulate and interpret statistical information? All too often the answer to these questions is ‘no’.”<sup>1</sup>

Twenty years later, and based on some very real changes going on in academic libraries, I think the answer is still “no”. Therefore, I would like to address the issue of whether the statistics we are collecting are of real use to us. Compiling is not the same as analyzing, and in order to analyze we have to have meaningful data. We are collecting quantitative information, but that information often does not translate into useful numbers for qualitative analysis. For example, in preparing our statistics we count the number of reference transactions, but this does not measure user satisfaction or whether the reference question was correct or provided the complete answer. Which leads me to one of my personal opinions of statistics. I think many of us may be familiar with the phrase, “lies, damn lies, and statistics” which has been paraphrased in various ways and attributed to various individuals. It is probably the only statement I remember from my one undergraduate

---

<sup>1</sup>Steve Hiller and James Self, “From Measurement to Management: Using Data Wisely for Planning and Decision-Making. *Library Trends* (Summer 2004): 129-155.

statistics course. It captures not only my own sometimes skeptical view, but perhaps the general viewpoint of many who are involved in the compilation of the annual numbers we contribute to national surveys. The *Library Trends* article also questioned how we "manipulate and interpret" statistical information. Let me share the following with you:

**“A mathematician, applied mathematician and a statistician all apply for the same job. At the interview they are asked the question, what is 1+1. The mathematician replies, ‘I can prove that it exists but not that it is unique.’ The applied mathematician after some thought replies, ‘the answer is approximately 1.99 with an error in the region of 0.01.’ The statistician steps outside the room, mulls it over for several minutes, and eventually in desperation returns and inquires, ‘so what do you want it to be?’<sup>2</sup>**

Too often when working with numbers and statistics we resort to the principle of “what do we want them to be” in order to make a case or prove a point. A possible representative example of that is the OCLC study of *Perceptions of Libraries and Information Resources* that was completed in 2005. They recently released a subset of that data entitled, *College Students’ Perceptions of Libraries and Information Resources*.<sup>3</sup> A number of facets of student use of libraries are reported, but there are three published statistics that I want to point out:

- 72 percent of college students begin their research with a search engine
- 2 percent of college students begin their search on the library website
- this survey is based on 396 students aged 15 to 57 and includes graduate students as well as undergraduates and is not limited to students in the United States

First, this is a very limited sample. Second, how well does it capture *our* campuses, many of which are residential, smaller institutions that might not have been represented in the sample? Third, the real issue is that the results of these small samples start to crop up as “facts” in academic articles,

---

<sup>2</sup> Gary C. Ramseyer’s First Internet Gallery on Statistics Jokes, <http://www.ilstu.edu/~gcramsey/Gallery.html> (Accessed March 14, 2007)

<sup>3</sup> Cathy DeRosa et al. *College Students’ Perceptions of Libraries and Information Resources*. (Dublin, Ohio: OCLC, 2006.) <http://www.oclc.org/reports/perceptionscollege.htm> (Accessed March 8, 2007)

and later in the popular press. A recent article in *Educause Quarterly* on students who are “net savvy” repeated the first two statistics<sup>4</sup> without clearly expressing the third which demonstrated that it was a limited survey sample. One could argue that this is a case of manipulating statistics to prove a point. But if the point of statistics is to have some comparisons, how would we use these statistics in our libraries? Numbers such as these may be used as benchmarks, but are not necessarily indicators of reality on our campuses. And benchmarks are not the same as measures that may be used to assess quality or performance. The statistics that we currently keep focus on inputs and outputs, but inputs and outputs are not the same as outcomes and impacts. As I’m speaking, please keep in mind that most of the numbers we collect, measure quantity, not quality.

I will also address how current changes in libraries are affecting the statistics we have been keeping and suggest that we need to change what we collect. We need to keep in mind what types of analysis needs to be done through these statistics. Numbers collected on their own have no meaning. It is only through analysis that statistics provide value and insights for decision-making. Most of what I’ll be referring to today are the traditional statistics that we report in various surveys such as the ALS. These reports collect numbers on budgets, collections, services, and staffing. I’d like to look at each of these areas and identify some of the challenges in the current surveys. Let’s consider first our budgets.

Most of us have dealt with shrinking allocations for resources, especially in the past five years. One measure of just how much our allocations may have shrunk is to consider the 1986 ACRL standard for academic libraries that stipulated that the library should receive 6% of the institutional education and general program budget. Perhaps I may be corrected on this point, but in all of the surveys that I have completed, I’ve never been asked for the dollars for the total institution budget. I have been asked to list our endowment. There isn’t any comparison point regarding what percent of the institution budget is allocated to the library. If we were to collect more detailed institutional budget and library allocation information, is that meaningful for peer comparison? Would that provide a more meaningful context for our reports to administrators? In some respects it may be a moot point because the current ACRL academic library standards have changed. The 6% ratio has been eliminated in later standards.

---

<sup>4</sup>George Lorenzo, Diana Oblinger, and Charles Dziuban, “How Choice, Co-Creation, and Culture are Changing What it Means to Be Net Savvy”, *Educause Quarterly*, 30:1 (2007): 6-12.

The current ACRL standards for academic libraries have a limited number of comparison points, or “input measures” and output measures.<sup>5</sup> This is not a complete list, but just a few of them:

Suggested Points of comparison: Input Measures

- Ratio of volumes to combined total student (undergraduate and graduate, if applicable) and faculty FTE.
- Ratio of volumes added per year to combined total student and faculty FTE.
- Ratio of material/information resource expenditure to combined total student and faculty FTE.
- Percent of total library budget expended in the following three categories:
  1. materials/information resources subdivided by print, microform, and electronic.
  2. staff resources, subdivided by librarians, full and part-time staff, and student assistant expenditures.
  3. All other operating expenses (e.g. network infrastructure, equipment).
- Ratio of number of students attending library instructional sessions to total number of students in specified target groups.

Suggested Points of Comparison: Output measures

- Ratio of circulation to combined student and faculty FTE.
- Ratio of interlibrary loan requests to combined student and faculty FTE.
- Ratio of interlibrary loan lending to borrowing.
- Interlibrary loan/document delivery borrowing turnaround time, fill rate, and unit cost.
- Interlibrary loan/document delivery lending turnaround time, fill rate, and unit cost.
- Ratio of reference questions (sample week) to combined student and faculty FTE.

The new standards have deviated from previous standards not only by the type of comparison data they are suggesting, but in purpose as well. The new standards are intended to provide a tool to assist libraries in:

- establishing individual goals within the context of their institutional goals

---

<sup>5</sup> Standards for Libraries in Higher Education, approved by ACRL Board in June 2004. <http://www.ala.org/ala/acrl/acrlstandards/standardslibraries.htm> (visited 3/14/2007)

- documenting the library's contribution to institutional effectiveness and student learning outcomes
- suggested points of comparison for peer and longitudinal comparison, *and encourage the development of other measures* [emphasis mine]

So, the standards have changed, but like many of our surveys the ACRL survey is still asking for the same outdated library figures. I like that “encourage the development of other measures” point, and will return to it—and to the standards in general—later. But first, in addition to the budget, let's consider another area we collect statistics on: our collections..

We all know our collections are changing. It is likely you're purchasing fewer print books, focusing on electronic content, and are leasing, not purchasing, much of your electronic content. Now, let's look at the standard questions on some of the statistical questionnaires. As per the ACRL standard of “ratio of volumes added per year”, the ALS asks for volumes added by purchase. If one is purchasing a subscription, or access to an electronic book, does this count as a volume? Are you really purchasing or are you leasing—and how is that counted? Do counts for number of books and number of journal titles measure something of value? For college libraries, is the size of our collection a measure of quality? If our mission is to support the curriculum, and our curriculum changes over time, wouldn't a serious weeding project be of more value?

Let us look at journals. Simply identifying the number of journals we subscribe to has become extremely complicated. If we carry a journal in both print and electronic format, do we count it once, twice, or even three times if we own a microformat? Identifying titles within aggregated services is also difficult—what does “subscription” mean when you have access through an aggregator? Trying to determine what we count as a journal title has become dependent on the detailed definitions provided by each survey. There is no standard or consistency in the journal title counts between surveys. Trying to address this complexity has resulted in some discussion in our review of the Oberlin stats. I questioned whether we could really capture a reliable count based on all the complexities, and further why did we need to? The answer I received points out the difference in how some of us use these statistics. One of my colleagues felt it was important to use the numbers to justify the budget costs for third party services such as Serials Solutions which are used to help manage our content. In his words, “having comparative data helps justify budget to outsource some of this” and that it is easier to

demonstrate if he can “show that the vast majority of my colleagues are facing the same kind of volume of things that need to be managed.” As I mentioned in the beginning of my talk, if one is using statistics to prepare reports for our administrations, each of us may use the same data in different ways.

Usage statistics for journals is another area of great concern. These are the types of internal statistics that may not be used for compilation for surveys, but are important for assessing our own operations. In a related complication, we’re experiencing more offers of “bundled” rather than individual journal title subscriptions. This leads to another area of concern in compiling statistics: accurate use counts. If any of you have been engaged in a use study for your periodicals, you’re probably aware of the variety of ways we measure use for print journals. Depending on your methods, you can get a pretty good idea of which journals never move off the shelf versus those that are never on the shelf. However, as we move to an electronic environment, the lack of standards for what constitutes “use” alters our ability to collect data. Without reliable, comparable use statistics, how can we make decisions regarding subscriptions for our collections? What percentage of use makes a “bundle” worthwhile, even if there are titles in the bundle that are never accessed?

At Macalester we are moving to an increased reliance on electronic journal collection, even with limited dollars. We see great value in electronic subscriptions, however we will have to find a way to measure and compare electronic use in order to be able to make responsible decisions when we have to choose between two titles—or two bundles. This is an area that will have an impact on staffing as we will have to figure out, first how we collect the data, and second, who will be analyzing it.

A third area related to our collections isn’t even being considered in our counts: institutional repositories. Consider the new role some of our academic libraries have taken on as publisher. I know situations vary, but many of you probably have various special collections that have been digitized, or maybe you’ve implemented an institutional repository and started to archive faculty publications in a digital format, or you are managing a digital collection of images. At Macalester we are collecting student Honor Papers in an electronic format and making them accessible via the web. There are, however, no current measurements of new content that the library has made available via means other than purchase. The current counts in ALS are for:

- Books, serial backfiles and other paper materials (include government documents)

- E-Books
- Microforms
- Audiovisual materials
- Current serial subscriptions
- Electronic reference sources and aggregation services

Aside from the obvious, another reason measures on repositories and library-as-publisher may be particularly relevant, is to document a transition that may be taking place in your own libraries. If you are ordering fewer books, how has this changed what your catalogers are doing? What are they spending their time on? Is it in the production of metadata for digital collections? Have responsibilities shifted in terms of managing the electronic licenses that are an added component to the workflow? Changes in the library aren't just limited to our collections.

I've spoken about budgets and collections and now I'd like to focus on service. The traditional collection of data on services has focused on transactions. For years we've been counting reference transactions. But what is the definition of a reference transaction now? Does it include electronic chat? Is the number of questions completed via email or a service such as Questionpoint included? What about text messaging, which some libraries are doing for reference? In the ALS definitions, a reference transaction is "an information contact that involves the knowledge, use, commendation, interpretation, or instruction in the use of one or more information sources by a member of the library staff" and include transactions "in person, by phone, by e-mail, by the Web". One particular issue I've had is the fact that we've always included a number count for transactions, but not the time spent on answering questions. I'm fairly certain that reference librarians at each of your institutions would confirm that there may be fewer questions, but the questions we get are more complex and take longer to complete. Yet we have no measurement of that particular trend, only anecdotal reporting. We see only that the number of transactions taking place at our reference desks are going down. Is this an accurate reflection of our reference services? What about the amount of time we are spending on instruction and integrating information literacy into the curriculum. We currently measure one aspect of instruction services, number of presentations. On the ALS survey they have at least included an item for the number of participants in those sessions, and they have included a series of questions related to the institution and information literacy. It's a step in the right direction.



Another area where service has changed in a somewhat dramatic way is in interlibrary loan. Again, situations may vary by institution, but for many of us our patrons may place their own holds or make their own, unmediated, requests to borrow materials not available in their libraries. How do we measure the impact of unmediated requests and the impact on circulation? Another question I have is that the standards for academic libraries asks for turnaround time for ILL requests, but they do not ask for the time it takes to process a book that is acquired for our collection. Some of your institutions may measure the time it takes on an average week for a book to be received in acquisitions and then processed by cataloging and finally available for circulation, but there is no national survey that requests that particular number. Is this a worthwhile measurement of comparison for our institutions? Does the size of staff in cataloging have an impact on the turnaround time? Might we compare book processing time and costs with ILL processing time and costs? Does the processing time for books have an impact on interlibrary loan requests? Again, I do not have an answer, but are these questions we should be exploring?

So, in addition to budgets, collections, services, what about staffing? The changes I've mentioned in collections and services, obviously, have an impact on staffing. On our surveys, we provide counts for our staff positions every year, but no-one asks us about *transitions* in our staffing. For example, at Macalester we were able to decrease our cataloging staff by one cataloger and moved some assignments around to create three different positions to address ongoing changes in the library related to creating digital content, managing a digital archives, and the need for more support for web services. There aren't any statistics that truly measure how staffing is changing in academic libraries as we move to digital collections and meet the needs of the 21<sup>st</sup> century.

Speaking of which, I want to return briefly to that OCLC statistic I mentioned early on in this talk: that only 2 percent of the students surveyed began their research at the library web site. With the increasing demands on our staff, is it an effective use of our time to put a lot of effort into the library web site? Two percent translates to about 36 students at Macalester. One could ask, how many students end up at the library web site after consulting a search engine? The same report stated that "48 percent ended up at a library Web site"<sup>8</sup>, but "only 10 percent of college students indicated that their library's collection fulfilled their information needs after accessing

---

<sup>8</sup> C. Rosa, op. cit., p.6-3

the library Web site from a search engine.”<sup>9</sup> If one were to consider this statistic reliable, it could be a rather depressing state of affairs, but it is another point to ponder in analyzing statistics as they apply to our own situations.

Finally, I want to address the issue of time spent collecting statistics. Some of the data is truly relevant for assessment of trends over time. We might question whether comparing the size of our collections to our peers over time has any true relevance. On the other hand, if we look at how the changing size and composition of our own collections over time has made an impact on our staffing, then maybe we can provide some relevant analysis to our administrations on the changing nature-- and value--of our libraries. Many of us share the standard statistics with our administrators, who then see that our gate counts are fluctuating, and our reference questions may be fewer. They don't see the changes that aren't reflected in traditional statistics. For example, we may provide 120 library instruction sessions, but that number doesn't immediately demonstrate that those sessions resulted in twenty-five one-hour consultations between librarians and students. Which leads me to the issue of “outcomes.”

The focus of our annual collection of numbers has traditionally been one of inputs and outputs: number of books in; number of books circulated, etc. If your campus is similar to mine, assessment is the new word of the day. The new ACRL standards take “outcomes” into consideration. Outcomes are defined as “the ways in which library users are changed as a result of their contact with the library’s resources and programs.”<sup>10</sup> Each of our institutions have similar missions, but we have different cultures and different relationships with faculty and administration. How then, do we design standard data collection that measures both comparative and individual impact on our institutions? How do we document and provide a rationale for the role of our libraries in the overall academic success of our students and the research production of our faculty? Is a measure of our impact the ability to achieve the 6% ratio of the overall institutional budget? Or is a measure of our impact the twenty-five students who received extensive one-on-one assistance with their research projects as a result of a one hour library instruction session? Is a measure of our value our contributions by providing interlibrary loan services for faculty research needs? While I don't have the answer to those questions, I do think the questions need consideration.

---

<sup>9</sup> Ibid, p. 6-2.

<sup>10</sup> ACRL Standards for Libraries in Higher Education.

It is time to start looking seriously at the kinds of numbers we collect and keep. Libraries have changed. We still do many of the same things we've always done, but we've taken on new formats, new roles, and new responsibilities. Our data collection needs to reflect this change, and needs to provide useful analysis for us to continue to grow and be relevant to our institutions. It helps to compare ourselves with peer institutions, but we also need to be able to show our administration, our faculty, our students, and ourselves, the impact these changes are having on our collections and services. What information is needed to assist in our decision-making that we don't currently have? Going back to my original set of questions, are the statistics used to provide analysis for:

- Process improvements?
- Strategic planning?
- Annual reports?
- Staffing adjustments?

To truly demonstrate and evaluate our role in enriching the experience of our faculty and students, we have to have more useful statistics. We need numbers that can be analyzed to provide a clear assessment of our impact. Remember that point in Standards for Libraries in Higher Education that I liked? We are *encouraged to develop new measures*.

On the other hand, maybe we don't need to keep all these statistics. If we want to assess how we are doing as it relates to our campus, perhaps we should follow the advice in a recent article in the December 2006 issue of *NextSpace* published by OCLC. In the article, "Are you Asking the Ultimate Question", we learn the ultimate question is "would you recommend us to a friend?"<sup>11</sup> Perhaps after every library instruction session, reference transaction or completed interlibrary loan request, we should ask our faculty and students: "would you recommend us to a friend?" Think about that. Unfortunately, the "Ultimate Question" may not do very well in measuring trends over time, so a combination of annual records of numerical transactions that have meaning, as well as surveys, and the development of new measures is in our best interests. I'm hoping that my comments today will lead to some suggestions on improvements that we may want to make in the statistics we compile and then analyze. We're encouraged to develop new measures, I say, let's do it.

---

<sup>11</sup> Tom Storey, "Are you asking the ultimate question?" *NextSpace*, (December 2006): 6-11.