Edward T. O'Neill, Ph.D.
OCLC Online Computer Library Center, Inc.

Lynn Silipigni Connaway, Ph.D.
OCLC Online Computer Library Center, Inc.

Timothy J. Dickey, Ph.D.
OCLC Online Computer Library Center, Inc.

Estimating the Audience Level for Library Resources

Note: This is a pre-print version of a paper published in *Journal of the American Society for Information Science and Technology*. Please cite the published version; a suggested citation appears below.

Abstract
WorldCat, OCLC's bibliographic database, identifies books and the libraries that hold them. The holdings provide detailed information about the type and number of libraries that have acquired the material. Using this information, it is possible to infer the type of audience for which the material is intended. A quantitative measure, the *audience level*, is derived from the types of libraries that have selected the resource. The audience level can be used to refine discovery, analyze collections, advise readers, and enhance reference services.

Introduction and Statement of the Problem

Current financial restrictions make it critical for librarians to use empirical data to assess and manage collections. Librarians assess collections to determine subject areas for acquisition, deaccession, digitization, preservation, and remote storage. They also must determine if the sources are relevant to their primary users' needs and expectations. One collection assessment method is to examine usage statistics, such as circulation and interlibrary loan data. Librarians employ usage data as one indicator of the materials' relevance.

Determining if the materials' content and presentation match the needs of the library's primary user groups is another form of collection assessment. The exponential amount of sources retrieved in the online environment can make it difficult to determine what content is appropriate for the intended audience's need. The audience level for a book theoretically represents the type of reader for which the resource is most appropriate, and thus can improve collection assessment and the development of a ranking system for discovery.

Estimating the audience level also can enhance information retrieval by increasing the relevance of items retrieved. Determining the audience level is difficult because there is no standard requiring the inclusion of this information in the bibliographic record, other than the Target Audience element in the MARC record and the Library of Congress Subject Heading (LCSH) form subdivisions which in terms of the audience level are both used primarily to identify juvenile material. The researchers hypothesized that the

audience level could be estimated from the types of libraries – research, academic, public, and school – that have acquired the resource.

WorldCat, OCLC's bibliographic database, serves not only as an aggregator of bibliographic data, but also includes detailed holdings information that can support such an analysis. In July 2007, the WorldCat database contained more than 81 million records and identified more than a billion holding locations for library resources. WorldCat includes a holding symbol for every member library holding an item represented in the WorldCat database. Each holding represents a discrete selection decision implying that the material is relevant to the library's patrons and is consistent with the library's collection development strategy. Thus, the totality of these individual decisions can serve as a indicator of audience level.

Literature Review

The literature on management and assessment of library collections is vast, but only recently has expanded to assess and describe collections by the characteristics of the libraries owning the collections. As early as 1979, Bonk and Magrill (pp. 305-313) attempted to collect an authoritative bibliography of the various methods for collection analysis. The principal methodologies at that time were either checklist-based, or based upon quantitative measures such as total volumes and total expenditures. Magrill's later, more exhaustive literature review of collection analysis methodologies revealed "variations on the traditional checking of standard bibliographies" (Magrill, 1985, 279). One of the most extensive bibliographies of collection assessment tools was that of Strohl (1999), who expanded the list of methods to include checklists, circulation data, citation analysis, the RLG-OCLC subject Conspectus, document delivery and ILL data, faculty

recommendations, and user-centered evaluation. Philips and Williams (2003) were able to add little to the literature in terms of assessment methods, though they documented that the number of studies had increased exponentially. They mentioned only one study which used WorldCat holdings data as an assessment tool (Senkevitch and Sweetland 1996, see below).

At the same time, researchers have suggested that the WorldCat database represents an "aggregate collection" (Lavoie, Connaway, & O'Neill April 2007, 107), which is appropriate for bibliometric study. Lavoie, Connaway, and O'Neill mined data from WorldCat to "map the landscape" of digital resources cataloged in WorldCat and held by member libraries, discovering more than one million digital resources within the database, and describing characteristics of this aggregate digital collection that support library decision-making. Bernstein (2005) studied a random sample of bibliographic records from WorldCat, in a "demographic study" to determine the characteristics of the aggregate monographic collection in the database (see also Schonfeld & Lavoie, 2006). These studies acknowledged that WorldCat does not "represent the totality of world library holdings" (Bernstein, 80), though as an aggregate collection, analysis of its contents "affords a high-level perspective on historical patterns, suggests future trends, and supplies useful intelligence with which to inform decision making" (Lavoie, Connaway, & O'Neill April 2007, 107).

Several researchers have specifically used WorldCat's holdings to evaluate various types of library materials in the aggregate collection. The findings of Perrault (2002), for instance, reinforce the applicability of the WorldCat database as an aggregate collection for research. Specifically, she reported that the presence and accuracy of

monographic holdings in WorldCat was mirrored by a profile of research libraries' collections. In two earlier studies, Perrault (1995; 1999) used the OCLC AMIGOS product as a source of data on general library collection patterns in the United States. Carpenter and Getz (1995), Ciliberti (1994), Velluci (in Gottlieb 1994; 1993), Gyeszky, Allen and Smith (1992), Harrell (1992), Joy (1992), Schwartz (1994), and Webster (1995) also used the OCLC AMIGOS Collection Analysis product for their research. However, many of these studies were evaluating the effectiveness of the AMIGOS tool itself rather than the collections themselves.

Connaway, O'Neill, and Prabha (2006) used WorldCat holdings specifically as the point of analysis, to identify a body of "last copies" and to provide data for deaccession, digitization, and preservation decisions. Other researchers used WorldCat holdings to assess collections. Serebnick (1992) identified and described small publishers' books owned by libraries and cataloged in WorldCat, while Serebnick and Cullars (1984) and Shaw (1991) assessed adult fiction collections included in WorldCat. The language and literature collections in WorldCat were identified and assessed by Sweetland and Christiansen (1997), while Wiberly (2002; 2004) focused on humanities and social science collections.

Researchers have attempted to exploit WorldCat's library holdings data as a generalized tool for library collection analysis. Wallace and Boyce (1988) offered an early example of bibliometric analysis of WorldCat holdings as a measure of journal "value." The authors determined that they could not support a solitary correlation between how widely a journal is held and the journal's score on other evaluative criteria, such as citation analysis, ISI impact factors, and circulation statistics. Senkevitch and

Sweetland (1996) adopted a similar approach to using WorldCat holdings as a verification tool for titles in an adult fiction collection. Their results exposed some discrepancies between a "standard" list and public library holdings of these titles in WorldCat. Budd (1991) used WorldCat holdings as a tool to evaluate library collections in comparison to a standard recommended core list of books, the *Books for College Libraries* (Association of College and Research Libraries, 1988). He tentatively was able to support this checklist based upon WorldCat holdings. Calhoun (2000) developed a general model for collection development which blends WorldCat holdings with two major sources of book reviews and the associated value of monograph publishers.

Several studies have used WorldCat holdings to measure the audience level of individual titles. White (1995), and later, Lesniaski (2004) used WorldCat holdings as part of a collection analysis tool for individual titles (see also Twiss, 2001). Both worked from the premise that the sheer number or paucity of libraries holding a title alone reflects its "difficulty effect" (White, 1995, 10); therefore, the most specialized research titles should have minimal worldwide library holdings, and the most basic and generalist titles, the maximum number of library holdings. The approach also assumed (with more empirical justification) that "libraries holding half or more of the items at a higher level certainly will hold half or more of the items at a lower level" (Lesniaski, 2004, 13). Bernstein (2005) also used WorldCat holdings to predict the level of an item, in his terms nonexistent, unique, scarce, or non-scarce. However, his analysis was based solely upon the number of libraries holding an item with the presumption that the more broadly an item is held, the more general its appeal and vice versa.

Algorithm

The current approach utilizes the knowledge acquired from the earlier studies and extends the difficulty effect described by White (1995) and further tested by Lesniaski (2004) and Bernstein (2005) by considering the types of libraries holding the resource. By assigning a weight to each type of library that owns the title in WorldCat, an audience level can be calculated for each title based on aggregate library holdings in WorldCat. This approach was originally reported by O'Neill (2003) and later described in more detail by Connaway, O'Neill, Prabha, and Snyder (2004).

An algorithm was developed to estimate the audience level for each WorldCat resource. The audience level is determined in two steps. First a weighted holdings value is derived, either using the target audience in the 008 field from the bibliographic record, or based on the types of libraries holding the resource. This weighted holdings value is a numeric value between zero and one. In the second step, the weighted holdings value is converted to a percentile to form the audience level.

If one of the following codes for the target audience had been assigned in the bibliographic record, the weighted holdings value for the resource was derived directly from the target audience code using the associated weighted holdings values:

0.00 **a** (Preschool),

0.10 **b** (Primary)

0.15 **c** (Pre-adolescent)

0.25 **d** (Adolescent)

0.15 **j** (Juvenile)

If none of the above codes for target audience had been assigned, the weighted holding value was calculated using a weighted sum based upon the types of libraries that hold the resource. The following weighting is used:

0.00 School libraries,

0.33 Public libraries,

0.67 Academic libraries,

1.00 Research libraries.

Research libraries are defined as those libraries who are members of the Association of Research Libraries (ARL) and academic libraries as those at non-ARL academic institutions.

Using the library holdings data attached to each item, the weighted holdings value is calculated for each resource in WorldCat. Only the four above types of libraries are considered when calculating the audience level. There are, of course, many other types of libraries among the OCLC members. However some of the other library types such as special libraries, government libraries, and library networks have very heterogeneous collections making it difficult to place them in the school to research library spectrum. Fortunately, these four types account for 93% of all WorldCat holdings, so excluding the other types of libraries does not have a major impact. The most significant impact of their exclusion is that there are a few resources for which the weighted holdings value can not be calculated. If a particular resource is only held by a special library, it will not have any useable holdings information; therefore, no weighted holdings value can be calculated.

As an example, *Build Community: the Leader's Guide to Building Community* (OCLC #65514085) is held by 12 OCLC member libraries as shown in Table 1.

Table 1
Computing Audience Level for
Build Community: the Leader's Guide to Building Community

Library Symbol	Library Name	Library Type	Weight
OUN	Ohio University	Research	1.00
KSU	Kent State University	Research	1.00
CIN	University of Cincinnati	Research	1.00
BGU	Bowling Green State University	Academic	0.67
TOL	University of Toledo	Academic	0.67
MIA	Miami University	Academic	0.67
HIR	Hiram College	Academic	0.67
YNG	Youngstown State University	Academic	0.67
OHI	State Library of Ohio	Other	x
OCO	Columbus Metropolian Library	Public	0.33
BGF	Firelands College	Academic	0.67
OSD	SEO Automation Consortium	Other	x

Four different types of libraries—research, academic, public and 'other' hold this book.

However, the other category is not included so these two libraries are ignored in

computing the weighted holdings value. The weighted holdings value for the book is

then:

$$\frac{\text{Sum of the weights}}{\text{Number of libraries}} = \frac{7.35}{10} = 0.735$$

Although the weighted holdings value is a valid measure of the audience level, its

meaning can be difficult to interpret. The distribution of the weighted holdings values for

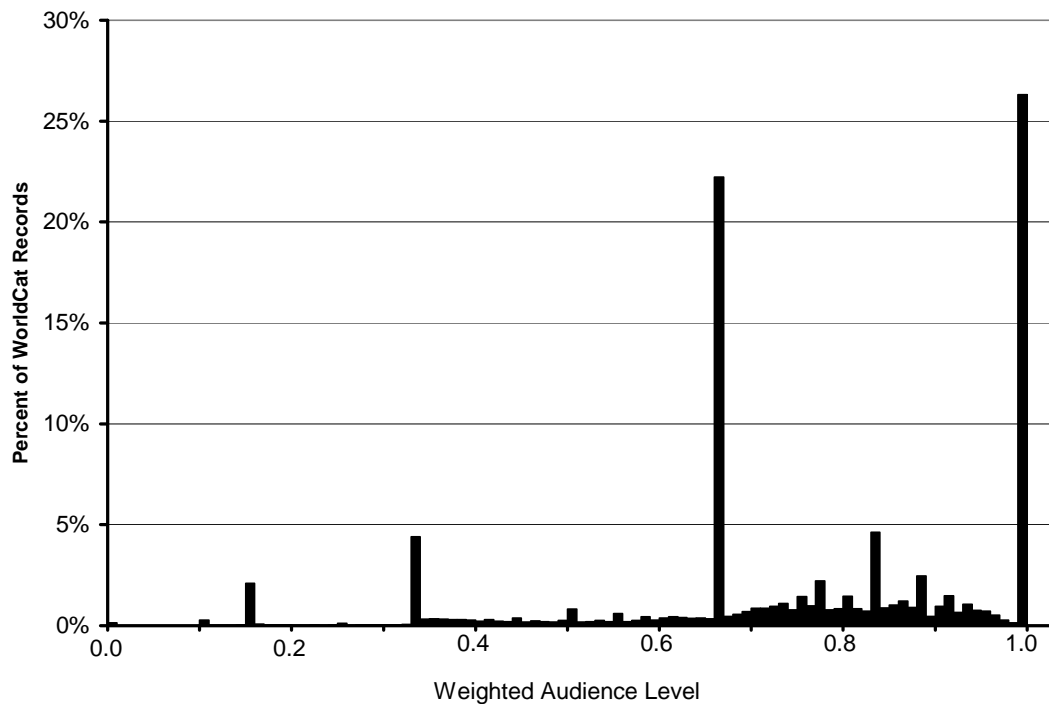all of the resources in WorldCat is shown in Figure 1.

Figure 1. Distribution of Weighted Holding Values

As can be seen, the weighted holdings values are not uniformly distributed and cluster at several points. The clustering observed at the lower values is primarily the result of using the target audience to derive the weighted holdings. The 'j' (juvenile) code is commonly assigned, creating a large cluster at 0.15. Similar, although much smaller, clusters are created by the other target audience codes. Approximately half of the resources in WorldCat are held by only a single library. These uniquely held resources generate a large cluster at 1.00 (resources held by a single research library), 0.67 (academic), 0.33 (public), and 0.00 (school). Smaller clusters also result from resources held by a small number of libraries. To simplify the measure and make it easier to interpret, the weighted holdings value is converted to a percentile to create the audience level. For the above example, the weighted holdings value of 0.735 is converted to a

percentile to form the audience level of 0.66. This audience level value indicates that 34% of the books in WorldCat have a higher audience level while 66% have a lower value.

The audience level is a property of the work rather than a property of a particular edition or manifestation and is computed for the work level as a whole.[1] In the above *Build Community* example, there is only a single manifestation of the work so this distinction was not relevant. This distinction is significant for works with multiple manifestations. The necessity of making this distinction was first observed for *Mother Goose*, the famous children's story. There are a large number of different manifestations of *Mother Goose* and some of the editions are rare and have very limited holdings while others are very widely held. Initially, when the audience level was derived at the manifestation level, it was observed that there was little consistency across editions; some editions had audience levels of 1.0, some had 0.0, and everything in between.

The rare editions are typically held by research libraries. Those editions held only by research libraries would receive a weighted holdings value of 1.0. It is these rare, or at least rarely held, editions that created the wide variation in audience level values for *Mother Goose*. Since audience level is a work property; the solution is to derive the audience level for the work as a whole and to use that value for all manifestations of the work. For computational efficiency, the weighted holdings is computed for each record (manifestation) in WorldCat and then combined to create the weighted holdings for the work.

---

[1] For the detailed definitions of work and manifestation as defined in for the Functional Requirements for Bibliographic Records (FRBR), see IFLA (1998).

To identify all the manifestations of a work, the audience level algorithm relies on the workset algorithm developed by Hickey and Toves (2005). Their algorithm has been used to "FRBRize" WorldCat. A second example illustrates the procedure for deriving audience level for *Courtroom Criminal Evidence*. This work has 4 manifestations as shown in Table 2.

Table 2

Audience Level Computation for *Courtroom Criminal Evidence*

OCLC No.	Total Holdings	Usable Holdings	Manifestation Audience Level
15504400	139	114	0.783825
29613712	161	117	0.769453
40393191	193	136	0.789426
62762763	174	124	0.758274

As with other works with multiple manifestations, the first step is to compute the weighted holdings values for each manifestation using the same methodology as in the previous example. The weighted holdings value is then calculated for the work as a whole by taking a weighted average of those for the individual manifestations. In this example, the weighted holdings value for the work is 0.775. The final step is the conversion of weighted holdings value to a percentile to create the audience level for the work, in this case a value of 0.76 (24% of the works in WorldCat have a higher weighted holdings value). By deriving the audience level value at the work level, the variability associated with rarely held and other atypical manifestations is minimized and the resulting audience level is more reflective of the content of the work as a whole.

Audience levels have been computed for each of the nearly 100 million individual resources in WorldCat and are beginning to be used in various OCLC applications. Access to the audience level values for WorldCat resources are made publicly available through the Audience Level Prototype.[2] It successfully was used to identify the scholarly books in WorldCat for Microsoft Live Search. Audience levels are also used in OCLC's FictionFinder prototype that provides access to 2.8 million works of fiction found in the WorldCat database.[3] It is also being used to enhance retrieval with the DeweyBrowser, and as an evaluation tool for the aggregate of works by and about an individual in WorldCat Identities.[4] It appears to be a valuable tool for analyzing and evaluating library collections and its potential in this area is being evaluated as part of the OhioLINK Collection Analysis Project.[5]

Evaluation of Algorithm and Findings

The testing of the calculations for various titles indicates the audience level is an accurate and appropriate measure. However, two test methodologies were developed and conducted to systematically evaluate the calculations.

The first test began with the generation of a random sample of 126 monographic titles held by an ARL library that was accessible to the research team. The team visited the library to examine some of the resources, which allowed researchers to determine if the audience levels were a meaningful measure of the target audience. The team examined the covers, title pages, table of contents, indexes, text, and images to assess the

---

[2] http://www.oclc.org/research/projects/audience/
[3] http://fictionfinder.oclc.org/; on this application, see also Pisanski & Žumer (2007).
[4] http://deweybrowser.oclc.org/ddcbrowser2/; http://orlabs.oclc.org/Identities/; on the DeweyBrowser, see also Vizine-Goetz & Mitchell (2006).
[5] http://platinum.ohiolink.edu/cbtf/oclcres.ppt

calculated audience level for each title in the sample. Digital images were captured for future reference and discussion. Although this evaluation was very subjective, it was encouraging to find that the audience levels appeared to be appropriate.

The second test compared the rankings of the audience level against ranking decisions made by human subjects. A sample of 30 books was ranked by each of 21 test participants, and a set of test rankings was created for each of the books. The test collection consisted of a stratified sample of 30 books from WorldCat. The books were all in the field of zoology, published in the year 2004, and representative of the entire spectrum of audience level rankings.

Zoology was chosen because it is a field with a wide variety of books ranging from children's books to highly specialized scholarly material. It also was believed that limiting the books to a single subject would facilitate comparison. The 2004 publication date was chosen since the acquisition and cataloging processes for 2004 books should be nearly complete but the books still would be current. All WorldCat records meeting the first two conditions were stratified by audience level (i.e. by their ranking within the entire database), and three were randomly selected from each decile.[6] Table 3 identifies the books in the sample.

---

[6] The ranking numbers used for data analysis differ slightlyfrom those used for the random sampling of the books. The sampling was done from Audience Level computations made on 18 January 2006; the following data analysis compares the test results to updated Audience Level figures, taking into account WorldCat holdings current to July 2007.

Table 3
Sample of Test Books

Audience Level Rank	Author	Title	OCLC record #	Audience Level	Average Subject Ranking	Holdings in WorldCat
1	Legg, G.	Octopusses and squid	54857977	0.06	1.60	213
2	Miller, H.	Mosquito	52757310	0.06	2.05	100
3	Burnie, D.	Bird [Eyewitness Books]	56189296	0.08	3.95	2914
4	Hall, D.	The ultimate guide to snakes and reptiles	56749767	0.14	13.05	83
5	Chittenden, R.	Birds of prey of the world	54718467	0.15	8.10	73
6	Mancini, J.	Guide to backyard birds	54415882	0.15	7.55	190
7	Romashko, S.	The complete collector's guide to shells and shelling	56960134	0.16	10.30	10
8		Curious critters of the natural world: Reptiles & amphibians	62674819	0.27	3.45	19
9	Haas, S.	Birds of Pennsylvania	60687811	0.31	11.50	17
10	Palmer, T.	Landscape with reptile: Rattlesnakes in an urban world	54046614	0.32	16.1	504
11	Thompson, B.	South Carolina bird-watching: A year-round guide	55700823	0.34	8.55	8
12	Patterson, B.	The lions of Tsavo: Exploring the legacy	54472084	0.34	14.65	347
13	Heinrich, B.	Bumblebee economics	56128472	0.37	22.85	1214
14	Humann, P.	Reef fish identification: Baja to Panama	56980668	0.42	14.40	74
15	Hartman, W.	A guide to the birds of Door County	57358137	0.46	11.95	2
16	Elzinga, R.	Fundamentals of entomology	50510931	0.51	21.25	1345
17	Gaston, A.	Seabirds: A natural history	56349814	0.51	18.75	361
18	Duff, A.	Mammals of the world: A checklist	56204329	0.57	19.40	355
19	Podulka, S.	Handbook of bird biology	57003728	0.60	21.55	303
20	Porter, R.	Birds of the Middle East	57148591	0.66	14.90	36
21	Bradley, R.	In Ohio's backyard: Spiders	57662538	0.66	10.60	20
22	Powler, C.	Dynamics of large mammal populations	57894103	0.69	25.90	441
23	Legros, G.	Fabre, poet of science	60576417	0.72	18.65	63
24	Fascione, N.	People and predators: from conflict to coexistence	54694499	0.74	20.65	249
25	National research council	Atlantic salmon in Maine	56493371	0.78	23.20	199
26	Borrow, N.	Birds of western Africa	57733231	0.80	13.95	91
27	Broughton, J.	Prehistoric human impacts on California birds	57203355	0.90	27.25	74
28	Barr, T.	A classification and checklist of the genus	55500979	0.91	25.55	20
29	Minter, L.	Atlas and red data book of the frogs of South Africa	61303229	0.93	24.00	5
30	Wallace and Dietz	Phylogeny and systematics of the treehopper subfamily	54460359	0.96	29.35	33

21 participants, who had no prior affiliation with the audience level project, agreed to test the results. All participants were volunteers from the staff at OCLC headquarters in Dublin, Ohio, representing a reasonably broad demographic spectrum within that population. Eleven females and ten males participated in the test, with job

responsibilities ranging from internships to upper management. Eleven participants held the Masters' degree in library or information science or its equivalent. The median educational level was two years' graduate study and the median age range reported was 40-50. Eight of the subjects reported professional cataloging experience, and four reported library reference experience.

The tests were conducted in the OCLC Usability Lab between July 17 and July 28, 2006. Each of the 21 participants was seated at a table with the 30 books arranged in a pseudorandom order; the order was identical for each participant. The instructions given to the subjects attempted to minimize any bias in the results by not dictating what criteria the participants should use or consider in their ranking. The participants were given the following instructions:

> "Please reorder these books in increasing order of difficulty, starting with
>
> pre-school books and proceeding to advanced scholarly material. Please
>
> let us know when you are done. Thank you for participating in this study."

The participants were given freedom to work however they desired in the space, and extra bookends were provided for their convenience. All but one of the participants produced a unique ordering of the books by perceived audience level.[7] Each participant's ranking of the books was recorded. With the exclusion of the one questionable data set, the tests produced a total of 20 valid sets of rankings.

None of the participants required the full ninety minutes allocated for their session. Each individual's approach may have been the greatest factor in the individuality

---

[7] One participant mistook the directions midway through the test, and returned the books to the original order. The subject re-took the test after debriefing; the data, however, were excluded from any further analysis.

of the results. Many participants worked from an initial "rough sort" into three or four piles, or as many as nine; others began by quickly identifying books at both ends of the ranking spectrum, and working inward. At least two participants took an extremely fast approach, scanning at most a few pages in the few books they opened; others took the time to read prefatory material and interior passages, sometimes even comparing passages from two or three works simultaneously. In post-test debriefings, most of the participants spoke of the simple presence or absence of features such as footnotes, bibliographies, charts and tables, or pictures. One participant claimed that the presence or absence of Latin genus/species names was a deciding factor in his choices.

Some individual books showed a greater variance in the test rankings than others. The two ends of the ranking spectrum were the most consistent since the ends are "bound" by the nature of the test, with no participant able to rank books below 1 or above 30. Figure 2 compares the subjects' rankings with the computed audience level rankings.
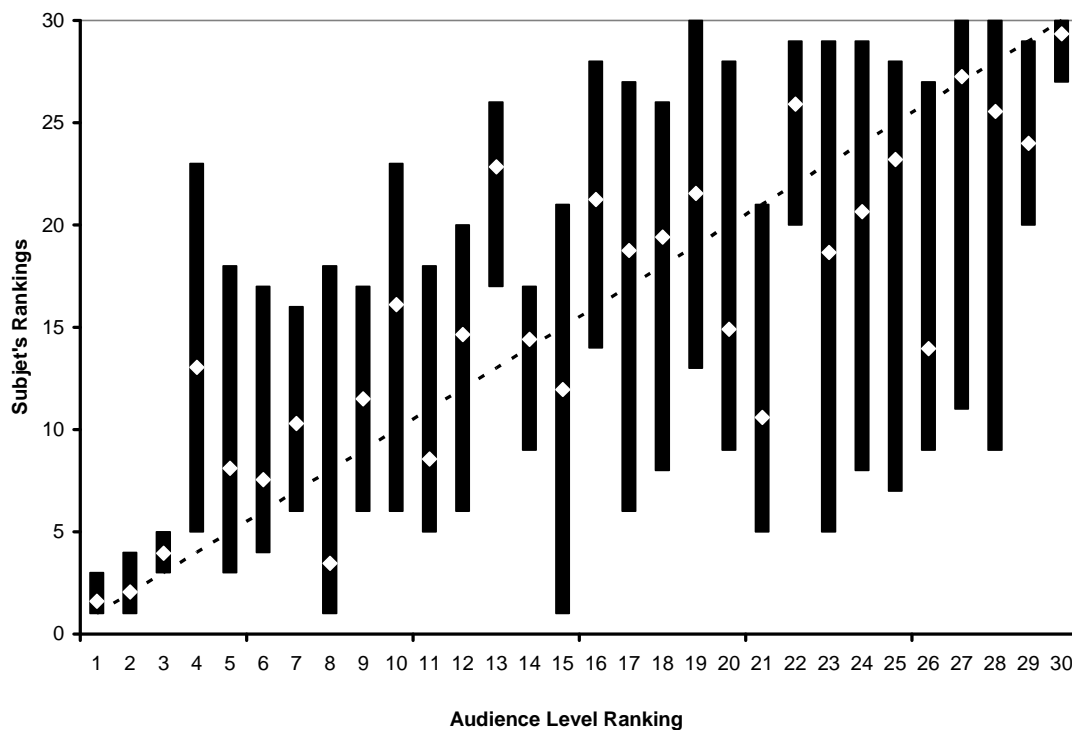
Figure 2. Audience Level vs. Subject's Rankings

The bars represent the range of the observed subject rankings and the diamond is the average of the subjects' rankings for the book. The dotted line is the ranking predicted by the audience level. As indicated by the length of the bars, there was wide variation in the subjects' assessment of the book's difficulty. Three books, identified by the wide range of observed rankings, seemed to be particularly difficult. Each presented to the subjects a different kind of cognitive challenge:

- *Fabre, Poet of Science* (Rank 23) is a reprint edition of a naturalist's diary, which could be considered either as light reading or as a scholarly work. The subjects ranked this book from a low of 5th to a high of 29th. The audience level for the book was high since nearly all of the libraries that hold this work and its two

earlier manifestations (1913 and 1921) are college and university libraries. In the case of many individual subjects, placement of this work was one of the last decisions made.

- *The Ultimate Guide to Snakes and Reptiles* (Rank 4) includes a great amount of information, but is a picture book. Its low audience level reflects the fact that the majority of libraries holding this work are public libraries.

- *A Guide to the Birds of Door County* (Rank 15) is a practical bird-watching guide, with hand-drawn pictures, which adds to the difficulty of assessing its level.

However, except for *The Ultimate Guide to Snakes and Reptiles*, even for these challenging books the average subject ranking was reasonably close to that predicted by audience level.

In all but two cases, the audience level ranking was within the range of the subjects' rankings. In the two cases where the level ranking was outside of the subjects' range, different factors may have contributed to divergence for these books:

- *The Ultimate Guide to Snakes and Reptiles* (Rank 4) discussed above, is an example of a discrepancy between the subjects' rankings and the audience level rankings. The subjects consistently ranked the book higher than predicted by the audience level.

- *Bumblebee Economics* (Rank 13) was also consistently ranked higher by the subjects than predicted by the audience level. This may be attributed to the subjects' (correct) perceptions of the book's basis in scholarly research and its

lengthy bibliography. They appeared to disregard the non-specialist, flowing

style of the author's prose.[8]

The Spearman Rank Coefficient of Correlation, a non-parametric statistical test,

was separately run on the results for each subject. The results with associated p-values

are shown in table 4. The Spearman Rank Coefficient test evaluates the degree of

correlation between the subjects' ranking and the audience level ranking. All the rho

values are significant at the 5% confidence level. The conclusion is that there exists a

significant correlation between the human subjects' ranking and the audience level

rankings. The most important result of this test is the indication that the audience level

and human subjects' perceptions are strongly correlated.

---

[8] *Bumblebee Economics* may be another installment in the trend for respected scholars to compose books specifically geared towards generalist audiences (*Fermat's Last Theorem*, *A Brief History of Time*, *Brunelleschi's Dome*). The first edition of *Bumblebee Economics* was cited in the *New York Times Book Review* as one of the "Best Books of 1979," (Nov. 25, 1979, section BR4), and it was nominated in both 1980 and 1982 for the American Book Awards.

Table 4.

Correlation between Audience Level and Tester's Rankings

Tester	Rho	P-value
1	0.6828	<0.0001
2	0.7526	<0.0001
3	0.8394	<0.0001
4	0.6512	0.0001
5	0.7531	<0.0001
6	0.7620	<0.0001
7	0.8363	<0.0001
8	0.5092	0.0041
9	0.5564	0.0014
10	0.8389	<0.0001
11	0.6966	<0.0001
12	0.7682	<0.0001
13	0.8469	<0.0001
14	0.7544	<0.0001
15	0.7237	<0.0001
16	0.7464	<0.0001
17	0.7918	<0.0001
18	0.5737	0.0009
19	0.8274	<0.0001
20	0.8852	<0.0001

Correlation between Audience Level and Holdings

As discussed earlier, both White (1995) and Lesniaski (2004) assumed that the number of libraries holding a title alone could be used to estimate its audience level or what White referred to as the "difficulty effect" (White, 1995, 10). Figure 3 depicts the relationship between the audience level and the average number of holdings.
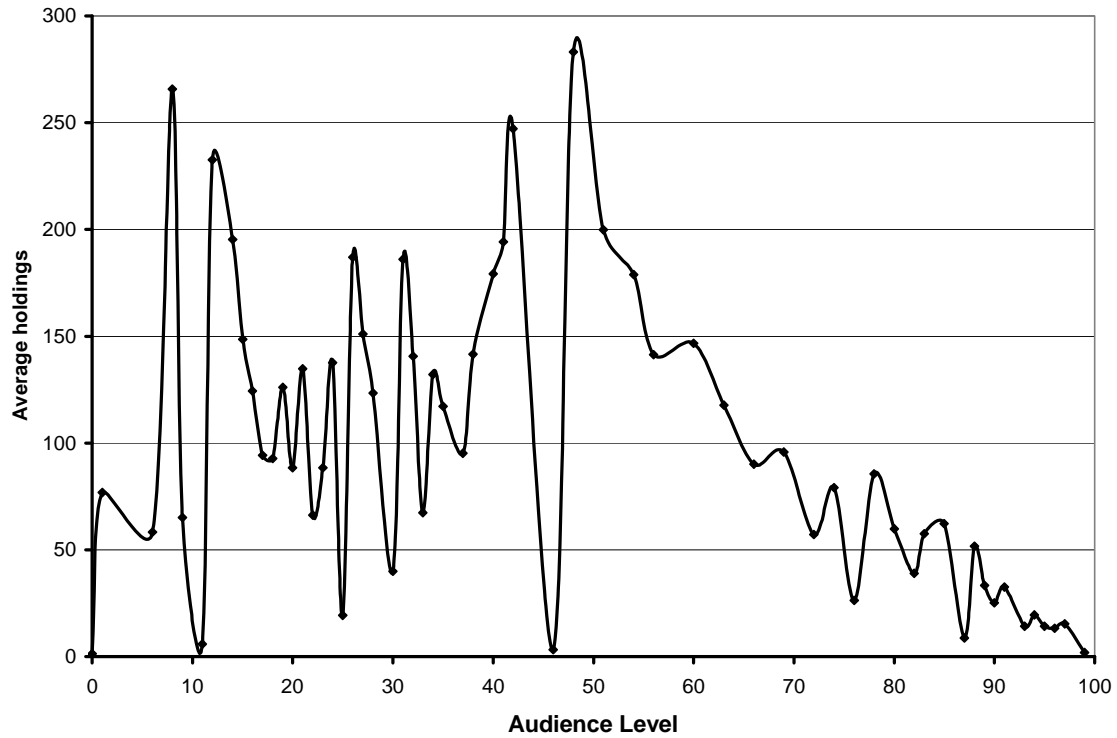
Figure 3. Relationship between Audience Level and Holdings

The figure indicates a strong inverse relationship between the audience level and the average number of holdings for audience levels greater than 0.5. Books with high audience levels are not widely held. However, the reverse is not generally true; books with low audience levels are not necessary widely held. Hence the number of libraries holding a resource is not by itself a good prediction of its audience level.

The number of holdings and the audience level, in fact, are measures of different although related attributes. The audience level is really a predictor of the target audience while within a given audience level, the number of holdings is a predictor of the perceived quality or popularity of the resource. Resources with very high audience levels by definition will be held predominately by research libraries. Since, compared to other

types of libraries, there are relatively few research libraries; resources with high audience levels never will be widely held. To be widely held, resources must have broad appeal.

Conclusions

The audience level is a valuable aid in identifying the appropriate resources for a particular audience. The algorithm produced audience level values that were consistent with those of human evaluators as demonstrated both by the analysis of the actual books and by the comparison of the algorithmic results to those of a test group of human subjects. Based on the findings of this research, the audience level is a new tool with the potential to improve information relevance for discovery and selection for collection analysis, readers' advisory, and reference services.

Since the audience level is a valid predictor of the target audience for a resource, it can be integrated into existing and new systems. The audience level has already been integrated into several OCLC prototypes - FictionFinder, WorldCat Identities, and DeweyBrowser - which can aid both librarians and users in discovering and selecting appropriate materials through various services. It currently is being applied to the OhioLINK Collection Analysis Project in anticipation of integration into future reference and collection assessment services. The audience level was used to enhance discovery of scholarly books in Microsoft Live Search; there is potential for integration of audience level into other discovery systems, such as WorldCat.org and WorldCat Local. This integration would benefit librarians and users in their discovery and selection of materials.

Acknowledgement

References

Amazon.com: Online shopping for electronics, apparel, books, computers, and more. (n.d.). Accessed September 20, 2006 from http://www.amazon.com.

Association of College and Research Libraries. (1988). *Books for college libraries: A core collection of 50,000 titles*. 3rd ed. Chicago: American Library Association.

Bernstein, J. H. (2006). From the ubiquitous to the nonexistent: A demographic study of OCLC WorldCat. *LRTS*, *50(2)*, 79-90.

Bonk, W. J., & Magrill, M. J. (1979). *Building library collections*. 5th ed. Metuchen, NJ: Scarecrow Press.

Budd, J. M. (1991, July). The utility of a recommended core list: An examination of *Books for College Libraries*, 3rd ed. *Journal of Academic Librarianship*, *17(3)*, 140-144.

Calhoun, J. C. (1998, January). Gauging the reception of *Choice* reviews through online union catalog holdings. *LRTS*, *42(1)*, 21-43.

Calhoun, J. C. (2001, July). Reviews, holdings and presses and publishers in academic library book acquisitions. *LRTS*, *45(3)*, 127-177.

Carpenter, D. E., & Getz, M. (1995). Evaluation of library resources in the field of economics: A case study. *Collection Management*, *20(1/2)*, 49-89.

Ciliberti, A. C. (1994, Winter). Collection evaluation and academic review: A pilot study using the OCLC/AMIGOS Collection Analysis CD. *Library Acquisitions: Practice and Theory*, *18(4)*, 431-445.

Connaway, L. S. (2007). Mountains, valleys, and pathways: Serials users' needs and steps to meet them. Part I: Preliminary analysis of focus group and semi-structured interviews at colleges and universities. *Serials Librarian*, *52(1/2)*, 223-236.

Connaway, L. S., O'Neill, E. T., & Prabha, C. (2006, July). Last copies: What's at risk? *College and Research Libraries*, *67(4)*, 370-379.

Connaway, L. S., O'Neill, E. T., Prabha, C., & Snyder, C. (2004, October). *Estimating audience level of monographs using holding patterns in WorldCat*. Paper presented at the 3rd annual Library Research Seminar, Kansas City, MO. Accessed September 20, 2007 from http://www.oclc.org/research/presentations/connaway/lrsIII_audience.ppt.

Gottlieb, J. (1994). *Collection assessment in music libraries*. MLA Technical Reports, 22. Canton, MA: Music Library Association.

Gyeszky, S., Allen, G., & Smith, C. R. (1992). Achieving academic excellence in higher education through improved library research collections: Using OCLC/AMIGOS Collection Analysis CD for collection building. In *American libraries: Achieving excellence in higher education*, 197-206. Chicago: American Library Association.

Harrell, J. (1992). Use of the OCLC/AMIGOS Collection Anlysis CD to determine comparative collection strength in English and American literature: A case study. *Technical Services Quarterly*, *9(3)*, 1-14.

Hernon, P. (1992). *Statistics: A component of the research process*. Norwood, NJ: Ablex.

Hickey, T., & Toves, J. (2005, April). FRBR work-set algorithm. Accessed Sept. 20, 2007 from http://www.oclc.org/research/projects/frbr/default.htm.

IFLA Committee on the Functional Requirements for Bibliographic Records. 1998. FRBR final report. Munich: K. G. Saur. Accessed Sept. 20, 2007, from http://www.ifla.org/VII/s13/wgfrbr/bibliography.htm.

Joy, A. H. (1992). The OCLC/AMIGOS Collection Analysis CD: A unique tool for collection evaluation and development. *Resource Sharing and Information Networks*, *8(1)*, 23-45.

Lavoie, B. F., Connaway, L. S., & O'Neill, E. T. (2007, April). Mapping WorldCat's digital landscape. *LRTS*, *51(2)*, 106-115.

Lesniaski, D. (2004). Evaluating collections: A discussion and extension of *Brief tests of collection strength*. *College & Undergraduate Libraries*, *11(1)*, 11-24.

Magrill, R. M. (1985). Evaluation by type of library. *Library Trends*, *33(3)*, 267-295.

OCLC Online Computer Library Center, Inc. *Audience Level prototype*. (2006a). Retrieved September 20, 2007 from http://www.oclc.org/research/researchworks/audience/default.htm#service#service.

OCLC Online Computer Library Center, Inc. *WorldCat.org*. (2006b). Retrieved September 20, 2007 from http://www.worldcat.org.

O'Neill, E. T. (2003). Estimating the audience level of books from holding patterns. Paper presented at the ASIST 2003 Annual Conference, Long Beach, California, October 22, 2003.

Perrault, A. M. (1995, Winter). The changing print resource base of academic libraries in the United States. *Journal of Education for Library and Information Science*, *36(4)*, 295-308.

Perrault, A. M. (1999). National collecting trends: collection analysis methods and findings. *Library & Information Science Research*, *21(1)*, 47-67.

Perrault, A. M. (2002). Global collective resources: A study of monographic bibliographic records in WorldCat. Retrieved September 20, 2007 from http://www.oclc.org/research/grants/reports/perrault/intro.pdf.

Phillips, L. L., & Williams, S. R. (2003). Collection development embraces the digital age: A review of the literature, 1997-2003. *LRTS*, *48(4)*, 273-299.

Pisanski, J., & Žumer, M. (2007). Functional requirements for bibliographic records: An investigation of two prototypes. *Program: Electronic Library and Information Systems*, *41(4)*, 400-417.

Schwartz, C. A. (1994, April). Empirical analysis of literature loss. *LRTS*, *38(2)*, 133-138.

Shonfeld, R. C., & Lavoie, B. F. (2006). Books without boundaries: A brief tour of the system-wide print book collection. *Journal of Electronic Publishing*, *9(2)*. Retrieved September 20, 2007 from http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;cc=jep;q1=Summer%202006;op2=and;op3=and;rgn=main;rgn1=citatio n;rgn2=title;rgn3=title;view=text;idno=3336451.0009.208;hi=0.

Senkevich, J. J., & Sweetland, J. H. (1996, Fall). Evaluating public library adult fiction: Can we define a core collection? *RQ*, *36(1)*, 103-117.

Serebnick, J. (1992, July). Selection and holdings of small publishers' books in OCLC libraries, a study of the influence of reviews, publishers, and vendors. *Library Quarterly*, *62(3)*, 259-294.

Serebnick, J., & Cullars, J. (1984). Find more like this: An analysis of reviews and library holdings of small publishers' books. *LRTS*, *28(1)*, 4-14.

Shaw, D. (1991). An analysis of the relationship between book reviews and fiction holdings in OCLC. *Library & Information Science Research*, *31(2)*, 147-154.

Sweetland, J. H., & Christiansen, P. G. (1997, March). Developing language and literature collections in academic libraries: A survey. *Journal of Academic Librarianship*, *23(2)*, 119-125.

Twiss, T. M. (2001). A validation of *Brief Tests of Collection Strength*. *Collection Management*, *25(3)*, 23-37.

Velluci, S. L. (1993, Summer). OCLC/AMIGOS Collection Analysis CD: Broadening the scope of use. *OCLC Systems and Services*, *9(2)*, 49-53.

Vizine-Goetz, D., & Mitchell, J. S. (2006). DeweyBrowser. *Cataloging & Classification Quarterly 42(3/4)*, 213-220.

Wagner, S. F. (1992). *Introduction to statistics*. New York: HarperCollins.

Wallace, D. P., & Boyce, B. R. (1989, January). Holdings as a measure of journal value. *Library & Information Science Research*, *11(1)*, 59-71.

Webster, M. G. (1995). Using the AMIGOS/OCLC Collection Analysis CD and student credit hour statistics to evaluate collection growth patters and potential demand. *Library Acquisitions: Practice and Theory*, *19(2)*, 197-210.

White, H. D. (1995). *Brief tests of collection strength: A methodology for all types of libraries*. Westport, CT: Greenwood Press.

Wiberley, S. E. (2002). The humanities: Who won the '90s in scholarly book publishing. *Portal: Libraries and the Academy*, *2(3)*, 357-374.

Wiberley, S. E. (2004). The social sciences: Who won the '90s in scholarly book publishing. *College & Research Libraries*, *65(6)*, 505-523.