WEEDING THE REFERENCE COLLECTION: A REVIEW OF THE LITERATURE

CHARLES D. BERNHOLZ

SCHOOL OF INFORMATION AND LIBRARY STUDIES
STATE UNIVERSITY OF NEW YORK AT BUFFALO

"Not buying new books so we don't have to weed old ones
is a dumb choice if we want people to use this library" (Ettelt, 1992, p. 259)

Weeding is an important tool in the quest for optimized collection development, and this is especially true for the reference collection. The history of, and the success derived from, the implementation of this procedure over the last one hundred years have, in general, been reported in a less than optimum manner, and only since Slote's (1989) methods have been generally accepted have the true gains become evident. This discussion examines the recent work in the area of weeding, with particular regard to the concerns of reference collections.

Effective operation of a reference desk or reference department in a library requires a combination of skills in the areas of personnel, reference materials, bibliographic instruction, and presentation. One area that always seems to be missed in such a list of requirements is the skills needed to effectively weed a reference collection.

While the idea of weeding has been around for almost one hundred years (Eliot, 1902), the underlying principles and the specific skills derived from these years of weeding experience have not yet elicited a clear set of guidelines. However, there are many offerings of weeding principles designed to help librarians manage their sites (e.g., Boll, 1980; Clark, 1991; Segal, 1980), even for those who plan to relocate (Kirby, 1995). This discussion will attempt to draw some of these materials together and to provide a glimpse of what has been learned about this task.

Slote (1989), in an important resource based on a quarter century of work in the area, has defined weeding as "removing the noncore collection from the primary collection area" (p. 274). While this may sound to be a fairly straightforward task, it will be seen that the actual mechanics of performing this activity vary widely and is the cause of much concern in libraries.

Furthermore, the actual decision process for selecting an item to be culled from the primary collection is not fixed, and varies according to the type of library under study and/or to the mandate of the library. Thus, the approach to weeding taken by a small public library has an outcome almost unique to small public libraries, which may differ substantially from the outcomes obtainable by special or by research libraries.

Slote (1989) demands that librarians approach this question with a very binary intent in mind. First, there will be items that will be identified as used, and second, there will be items that will be identified as unused. The methods that he proposes require that these two complementary subsets be ascertained accurately. Following this stage, the decisions to weed have already been made for the librarian: "no volume in the core collection should be considered for weeding . . . [a]ll books in the noncore collection are candidates for weeding, and probably should be weeded" (p. 83).

Ettelt (1992) asks the five pivotal questions that all librarians must ask at one time or another when their library reaches the point of capacity:

a) What percentage of our books *ever* gets used?
b) What percentage gets used in one year?
c) What subject areas get used and how much?
d) Why do we have to keep buying so many new books? Can't we cut back?
e) How likely are we to be wrong in weeding (or do we need a bigger library)? (pp. 257-259)

By answering these questions, and considering the repercussions of not weeding, libraries may determine

their best resource mix for their patron base, and move toward a more concentrated collection. Slote (1989) suggests that "weeding seems to increase accessibility, improve efficiency, reduce costs, and in many other ways improve collections and services to the average user" (p. 27).

Slote (1989) points out two very important criteria to keep in mind whenever the question of weeding is discussed. First, "librarians must use considerable judgment, experience, and professionalism" (p. 44) in determining the best use thresholds for weeding. This statement suggests that there can be no *fixed* rule to apply to all situations. Second, the advent of increased computerization in library systems has come to the aid of librarians who fear to weed. The advantages of online catalogs are widely discussed (Duval & Main, 1992; Matthews, 1986) and one of the many positive byproducts of their implementation is that very high levels of use may be set to use as a weeding criterion. Slote (1989) has noted that members of well-integrated automated library systems may always fall back on their neighbors to supply an item that has been weeded from their own library's collection, but which remains in the neighbor's collection because of differences in their library's usage. As a result, criterion of 90 to 95 percent use can be imposed, which would remove a substantial number of items from any library.

At one end of the weeding procedure option scale, the circulation statistics available from a large OPAC system can provide the necessary use data for culling the collection. Each item which has not circulated in some fixed time period as identified by the circulation summary statistics from the system is automatically a candidate for weeding. With these statistics available at any point in time from the onset of the sampling, different use thresholds may be set and subsets of weeding candidates may be calculated. This allows for the very high (90 to 95%) use criteria that Slote (1989) discusses for members of an online union catalog.

In fact, this can also serve as the basis of a very inexpensive and conservative model at the other end of the option scale. In this case, the spine of each book is marked to identify an item that has been used. Once marked, an item need not be remarked. At some fixed threshold of use, calculated by simple count data of the marked holdings, all those books so far unmarked are automatically weeded from the collection. If necessary, the time to a fixed threshold can be readjusted so that at that moment a subset of non-core, unmarked volumes is known. The advantage of this weeding procedure is that it does not require an OPAC support system and the statistical requirements are very modest (Slote, 1989).

But there are potential shortcomings in these procedures, particularly in regards to spine marking. Slote (1989) has noted that an accurate weeding signal requires between one and five or more years to develop. This is due to the relative slow rate at which many items are used, and an early termination of the sampling would suggest that a very large proportion of items be culled. Lee (1993)—in a mathematical analysis of non-action, item relegation to storage, and weeding—reminds librarians that circulation statistics always show random variation, so that the mean value for an item's circulation rate (e.g., three times per year) does not mean that the item circulates exactly three times each year but only achieves this mean rate over time. This variation clearly will influence the rate at which items can be marked—even only once—and only by setting use thresholds high will an adequate amount of time pass (1 to 5 years according to Slote). The only way to offset this wait is to reduce the use threshold, but by doing so the library runs the risk of weeding an item that, due to random variation, was not circulated during the time period of the investigation.

Lee's use of the Poisson distribution highlights this point.[1] In Lee's investigation of a library setting, if an item has an historical mean circulation of once per year, the probability that the item will *not* circulate in the next year is 36.8%. For an item with an historical mean circulation of three per year, the probability that it will *not* circulate in the following year goes down to only 5%. At two years, the prediction of future demand for this distribution goes down substantially. For the former once-per-year case, the probability that it will not circulate is only 13.5%, while for the latter three-times-per-year example the probability of not going into circulation is 0.25% (pp. 148-50; see especially Table 1 on p. 150). Thus, it is clear that the duration of the sampling can have a major influence on the findings, *especially for those items that are not, on average, frequently circulated*—i.e., those items with low historical mean circulation. But these are precisely the items that weeding is supposed to identify, so Slote's (1989) demand that "[a]ll books in the noncore collection are candidates for weeding, and probably should be weeded" (p. 83) is supported by these mathematical results.

This randomness and the potential penalties associated with it are also reinforced by the observations that, very frequently, very large proportions of the collection do not ever circulate. Slote (1989) reported on a University of Pittsburgh study that indicated that

39.8% of all the books that the university library had bought were never used during the first six years in the library (pp. 69-70, 83). This is precisely the reasoning behind Ettelt's (1992) initial question—what percentage of our books is ever used?

Clearly, even with this information, the only result is depressing support for the notion that library funds have been wasted in acquiring books that never will be circulated and that "weeding attacks the problem after the fact" (Slote, 1989, p. 83). Statistical models such as Lee's (1993) and Egghe's (1993), which uses the 80/20 percent rule for directing acquisition and weeding, are always *post hoc* methods.[2]

Finally, it makes sense—in terms of raw dollars—to weed. Joswick and Stierman (1993) have returned to a point made by Slote (1989)—it is extremely expensive, in terms of library construction costs, to house a single book. Using Slote's equation and more up to date (1991) construction figures, Joswick and Stierman (1993) estimated that the cost of housing a single volume in a new academic library is between $12 and $95, with a mean cost of approximately $39 (p. 114). Slote's (1989) findings indicated that, even with his 1987 data, the total cost of adding new facilities to a present library rises substantially because the design of the building must also incorporate facilities for non-book space as well. Indeed, he suggested that the mean increase in space expansion necessary to house each additional volume is 40 times the amount of space occupied by the book itself (p. 4). Thus, weeding makes economic sense, if only to defer any construction plans into the future.

Lee's (1993) article provides calculations which take into account the relative costs for storage, for ILL, and for users who need to wait for retrieval of materials that have been removed from circulation. All indicate that weeding is an important way to reduce costs for libraries, and that stashing low usage materials in storage, let alone keeping everything, is just not a viable alternative.

Such worries about physical cost can be substantially reduced by modifying the collection itself, and there are a number of reports on the impact of, for example, CD-ROMs on the policy and implementation of weeding (e.g., Truett, 1990, pp. 64-67).

But what about reference collections? Are not these subsets somewhat special and do not these items, one might think, possess an unusually high mean usage in comparison to the rest of the collection? Bopp and Smith (1995) have described the characteristics of reference collections and note that the "neglect of weeding can impair the efficiency of reference services" (p. 294), pre-

cisely because a large component of this specific collection is made up of serials, almanacs, encyclopedias, and yearbooks that are extremely time-sensitive. In other words, one might conclude that the specific characteristics of some of the primary reference materials themselves should *automatically* stimulate a level of weeding.

In a very important collection of articles on weeding, entitled *Weeding and maintenance of reference collections* (Pierce, 1990), all facets of the mechanics of weeding and maintenance of reference collections are examined. Specific examples of weeding in reference collections are presented and, regarding the question of assumed high usage, it is interesting to note a study at the University of Wisconsin-Eau Claire library (Engeldinger, 1990). In an examination of the reference collection that set the critical usage threshold of one use in a five year period, one-third of this library's titles could be culled. If the threshold was set to at least two uses in the same period—i.e., if a more reasonable level of expected use was used—*over half* of the reference titles could be removed. Besides the sinking feeling that these statistics present, they nonetheless support the very substantial need for weeding, and show that those libraries that do not have a formal policy on weeding are limiting their success and depleting their budget (Engeldinger, 1986). Studies at UCLA (Waters, 1990) and at Virginia Commonwealth University (Clark & Cary, 1995) have demonstrated the need for very careful analysis of personnel, collection development, acquisition, and weeding for direction into the next century. Waters (1990) describes her presentation in terms of "why I believe we are stronger for having survived" (p. 93) the realities of tight budgets of the 1980s, and their direct influence on the 1990s and beyond. The proactive efforts at UCLA included a very focused weeding committee.

Perhaps the most important paper in this collection concerns a study conducted at Iowa State University (Mathews & Tyckoson, 1990). Within this presentation, weeding is shown to have several important effects.

First, weeding immediately reduces the clutter on the shelves, and the removal of underused items improves circulation statistics because the remaining items are both easier to find and, in most cases, newer. Both of these improvements have been observed in other studies (e.g., the results obtained in Nova Scotia and reported in Slote, 1989, pp. 70-71).

Second, a number of reasons why libraries have not weeded are cited. Other, more interesting, professional responsibilities; the substantial time investment required

by a weeding review; accreditation standards that discourage reducing the collection; and general aversion to discard or store materials can all contribute to putting off a weeding program (Mathews & Tyckoson, 1990, p. 130). An examination of these can provide substantial reason for the reluctance of libraries to participate in this activity, and this does not include the inhibition imparted by increased costs, dwindling budgets, reduced space, or fewer personnel (Waters, 1990). The implicit message here is that the psychology behind much of this resistance must be changed in the future if well-run library programs are to survive.

Third, the balance between the fundamental characteristics of a reference collection and that of a general library ensemble is discussed, and important studies, including the one by Engeldinger (1986), are presented. This historical background *alone* makes this article a crucial one for any examination of the weeding literature. In this subset, the two philosophical approaches—keep all works regarding reference materials in a reference collection, and the primary use of reference materials should be to provide support for reference librarians and for their use—are shown to be, in most cases, not diametrically distinct. In the former case, large collections will be, by definition, the outcome of the application of this philosophy, where amassing a comprehensive collection is the goal. For the latter example, there is a transition in the site of the item. If the reference desk librarians find that they no longer use a specific item, then this item is moved from the private domain of the reference collection to the general, circulating one. In this specific case, the impact of weeding is even more relevant, because the actual penalty of using outdated material—providing incorrect answers to inquiries, using obsolete materials—is extremely high for the reference area.

Reallocation of materials into the general collection accomplishes two things. First, the general anxiety against culling collections is reduced because the item has not been discarded but only moved elsewhere within the system. Second, a very sharp weeding discipline converges to a non-circulating reference collection that exhibits zero growth. The Iowa State University library found that their best results were obtained if they attempted to apply this second philosophical approach to their reference collection.

The conclusions reached as a result of implementing this procedure point to a need for librarians to remain constantly vigilant to the usage of the materials in their collection. With the arrival of more and more vehicles to supply more and more information, any at-

tempt to cull to stay in a zero growth position requires an extreme amount of resource monitoring. Here again, the statistics of usage—and Ettelt's basic questions, for example—are crucial. Courage and a very strong understanding about the materials within the collection are two additional requirements to weed to the zero growth standard.

Finally, Mathews and Tyckoson (1990) provide a systematic approach to weeding that might work in other environments, and which illustrates the procedures installed at Iowa State. In that discussion, some very important procedures are introduced, including a plan to integrate various members of the weeding team in a manner that provides the best psychological climate for discussing and implementing weeding. For example, this policy at Iowa State was developed so that "the political benefits of such a program far outweigh the occasional change in disposition that a bibliographer may make over the reference librarians' suggestions" (p. 142). Such forethought can be particularly important during times of tight finances where tension among staff may be elevated, and this consideration has been commented on by other authors (e.g., Vincent, 1990).

Are there any down sides or limitations to these procedures for weeding reference collections? Perhaps there is one that requires special attention.

It was noted earlier that the technological appearance of OPACs has changed library collection statistics in a substantial manner, and now inventory statistics are easily available. This should in turn lead to easier decisions regarding collection weeding in the libraries with such systems, but when one wants to attack the reference collection with the same intent, the computerized statistics are absent because there are no circulation statistics available for reference materials that do not leave the library. This necessitates falling back on the Slote spine marking procedure, for example, and illustrates a restriction imposed on weeding this specific area of the collection. It will be important for librarians to develop methods to expedite the marking of reference materials so that this section of the library may join other sections that have the computerized support that the general collection circulation statistics can afford. Without accurate usage statistics, the underlying procedure for weeding the reference collection will fail.

Nevertheless, the penalty for failing to implement weeding is extremely high. Evans (1995) concisely stated three options that full libraries face: construct a new library; split the full one into two or more smaller facilities; or weed. The last opportunity is the only one which can be accomplished without the substantial cost

of additional space, and in times of reduced budgeting this may be the *only* alternative. It is clear that the financial aspects of library management have now been changed, and that the need to use resources—*all* resources—more effectively dictates that libraries reconsider the use of weeding as a practical collection development device. The mathematical models of circulation and of weeding facilitate its use and—along with the growth in the documented weeding experiences of different types of libraries—provide both instruction and inspiration for applying this tool. The cost effectiveness of this mechanism may in the end be the most prominent stimulus to its use.

Perhaps Trueswell (1965), in discussing his findings, presented the best reasons for weeding:

> The effective result of this approach is that there may be a quantitative way of maintaining the holdings of certain libraries at a reasonable level and a way of reducing the disappointment of the user who is looking for a specific volume. (p. 25)

These sound like important goals for all librarians today.

NOTES

1 The Poisson distribution is used to describe discrete random variables (Levin, 1987, pp. 229-34). It has been successful in characterizing the observed distributions of everything from the number of men in the Prussian Cavalry Corps killed by horse kick (Gerlough & Barnes, 1971, p. 6), to the number of ticks on sheep, or of telephone calls placed, or of photons reaching the retina, or of words misread in a text (Haight, 1967, pp. 100-107).

2 The 80/20 rule is of special interest to librarians since the 1960's when the ratio was used to identify those 20% of a collection's materials which were generating 80% of the circulation. This is an application of Pareto's "Law," a 19th century declaration regarding income distribution, and the relationship between production volume and the distribution of wealth (Cirillo, 1979, pp. 61-87, offers a useful review of this early excursion into econometrics). More recently, the prospect of such metrics stimulated thought on library efficiency, especially after demonstrations of computing for libraries a level of 99% of circulation requirements from only 25 to 40% of the current library holdings (Trueswell, 1965, p. 24). Turner (1980) provides an interesting discussion of Trueswell's work.

REFERENCES

Boll, J. J. (1980). *To grow or not to grow?: A review of alternatives to new academic library buildings* (Library Journal Special Report #15). New York: R.R. Bowker Co.

Bopp, R. E., & Smith, L. C. (1995). *Reference and information services: An introduction* (2d ed.). Englewood, CO: Libraries Unlimited.

Cirillo, R. (1979). *The economics of Vilfredo Pareto.* London: Frank Cass.

Clark, L. (Ed.). (1991). *Guide to review of library collections: Preservation, storage, and withdrawal.* Chicago: American Library Association.

Clark, J. M., & Cary, K. (1995). An approach to the evaluation of ready reference collections. *Reference Services Review, 23*(1), 39-43.

Duval, B. K., & Main, L. (1992). *Automated library systems: A librarian's guide and teaching manual.* Westport, CT: Meckler.

Egghe, L. (1993). Exact probabilistic and mathematical proofs of the relation between the mean m and the generalized 80/20 rule. *Journal of the American Society for Information Science, 44*(7), 369-375.

Eliot, C. W. (1902). The division of a library into books in use, and books not in use, with different storage methods for the two classes of books. *Library Journal, 27*(7), 51-56.

Engeldinger, E. A. (1986). Weeding of academic library reference collections: A survey of current use. *RQ, 25*(3), 366-371.

Engeldinger, E. A. (1990). "Use" as a criterion for the weeding of reference collections: A review and case study. In S. J. Pierce (Ed.), *Weeding and maintenance of reference collections* (pp. 119-128). New York: Haworth Press.

Ettelt, H. (1992). Accountability in book acquisition and weeding. *Reference Librarian, 38*, 257-259.

Evans, G. E. (1995). *Developing library and information center collections* (3d ed.). Englewood, CO: Libraries Unlimited.

Gerlough, D. L., & Barnes, F. C. (1971). The Poisson and other probability distributions in highway traffic. In *Poisson and other distributions in traffic.* Saugatuck, CT: Eno Foundation for Transportation.

Haight, F. A. (1967). *Handbook of the Poisson distribution.* New York: John Wiley.

Joswick, K. E., & Stierman, J. P. (1993). Systematic reference weeding: A workable model. *Collection Management, 18*(½), 103-115.

Kirby, L. (1995). Door to door: How to get your library moving . . . painlessly. *School Library Journal, 41*(2), 26-27.

Lee, H.-L. (1993). The library space problem, future demand, and collection control. *Library Resources and Technical Services, 37*(2), 147-166.

Levin, R. I. (1987). *Statistics for management* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Mathews, E., & Tyckoson, D. A. (1990). A program for the systematic weeding of the reference collection. In S. J. Pierce (Ed.), *Weeding and maintenance of reference collections* (pp. 129-143). New York: Haworth Press.

Matthews, J. R. (Ed.). (1986). *The impact of online catalogs.* New York: Neal-Schuman Publishers.

Pierce, S. J. (Ed.). (1990). *Weeding and maintenance of reference collections.* New York: Haworth Press.

Segal, J. P. (1980). *Evaluating and weeding collections in small and medium-sized public libraries: The CREW method.* Chicago: American Library Association.

Slote, S. J. (1989). *Weeding library collections: Library weeding methods* (3rd ed.). Englewood, CO: Libraries Unlimited.

Trueswell, R. W. (1965). A quantitative measure of user circulation requirements and its possible effect on stack thinning and multiple copy determination. *American Documentation, 16*(1), 20-25.

Truett, C. (1990). Weeding and evaluating the reference collection: A study of policies and practices in academic and public libraries. In S. J. Pierce (Ed.), *Weeding and maintenance of reference collections* (pp. 53-68). New York: Haworth Press.

Turner, S. J. (1980). Trueswell's weeding technique: The facts. *College & Research Libraries, 41*(2), 134-138.

Vincent, S. F. (1990). Let's get rid of it: A reference librarian's battle cry. In S. J. Pierce (Ed.), *Weeding and maintenance of*

reference collections (pp. 149-151). New York: Haworth Press.

Waters, M. B. (1990). Client-driven reference collections for the 1990s. In S. J. Pierce (Ed.), *Weeding and maintenance of reference collections* (pp. 93-102). New York: Haworth Press.