

and substantial scholarly constituency. This makes the Newberry's case, warts and all, fundamentally different from that of the New York Historical Society. Needless to add, the same could be said of any of the other four libraries in the study.

One of the things that sets America so radically apart from Europe is the way in which our cultural patrimony is distributed among so many independent libraries and museums. It is arguable that such a decentralized approach to preserving and making accessible the past is preferable to an overly controlled, overly centralized approach. From the perspective of one who has spent his professional career within the walls of large research universities, I can only admire the ways in which these libraries have served to complement the work of the academy, through both their collections and their programs. It would be hard to imagine the pursuit of historical and humanistic scholarship without them. Anyone who cares about them, indeed anyone who is concerned about the future of nonprofits in general, should pick up a copy of this book. At the very least, it should be required reading for all trustees and officers of institutions. I hope that Mr. Bowen keeps his word and that the Mellon Foundation sponsors future case studies as readable, as provocative, and as useful as this one.—*Michael Ryan, University of Pennsylvania, Philadelphia.*

White, Howard D. *Brief Tests of Collection Strength: A Methodology for All Types of Libraries.* Westport, Conn.: Greenwood Pr., 1995. 208p. \$55. (ISBN 0-313-29753-3).

Librarians have long quested for resource sharing and cooperative collection development. The goal seemed near when, in the early 1980s, the Research Libraries Group promulgated the *Conspectus* as an instrument through which all libraries could use common categories and a common language to describe their holdings. But the expected cooperative rewards never materialized. Libraries' inconsis-

tent self-assessments figure prominently in postmortem explanations. *Conspectus* rankings have been highly subjective, and the "verification studies" that would calibrate scores across institutions have proved both difficult to prepare and expensive to implement.

Howard White has probed the evaluation dilemma for more than a decade. This book offers his solution: "a new, relatively brief test to assign libraries a score for existing collection strength in a subject area." Each "brief test" consists of forty titles, divided evenly among ten-item segments that correspond to the *Conspectus*' four collection levels ("minimal coverage," "basic coverage," "instructional collections," and "research collections"). More than three hundred sample tests, for the most part constructed and applied by White's library school students, reveal a cheap and simple approach that provides reasonably consistent results. The sample tests also evince a methodologically satisfying pattern in which a library holding more than half the test items for any particular *Conspectus* level will own that many or more items from all of the lower levels. The tests thus bear out the hypothesis that real-life collections do not combine weak holdings of basic works with a strong representation of the esoteric. A final wrinkle validates the *Conspectus* level to which test creators assign each sample title—initially a subjective process—by tallying that title's holdings on OCLC. Although many libraries own the test items associated with "basic" collections, titles that test for "research" collections are held only sparsely.

As White himself acknowledges, this innovative approach invites methodological disputation. For instance, though this short volume is blessedly free of mathematical jargon, we are given neither empirical nor statistical arguments to justify fully the choice of forty items. The author eloquently defends testing economy and common sense, but does not explain why tests with ten items for each of four Con-

spectus levels are better than instruments containing thirty-two (or sixty, or eighty) items. (He does at several points suggest that larger tests would correct for some occasionally incongruous results.)

Although recognizing the distortions that errors in searching or counting can produce when each test category includes but ten items, White counsels readers to attend to his message rather than haggle over specifics. Some matters of seeming detail may nonetheless be significant. The test bibliography for American studies, for instance, lists the *Hispanic American Periodicals Index (HAPI)* as a sparsely held and hence research-level title. *HAPI*, one of two core indexes in Latin American studies, is an unlikely choice for a test in American studies. More important, it is a title very broadly held within the context of Latin American studies. White indicates that as many as 150 libraries can hold a title that will test for a research collection, his most restrictive category. But the *HAPI* example suggests that his definition would relegate virtually the entire literature of Latin American studies to this rarified niche. As White acknowledges, we need further study of how publications universes and collection sizes vary among fields.

These and similar questions suggest issues to clarify and refine. The book itself identifies other areas for additional work. For example, more inclusive bibliographic databases and increasingly sophisticated computer capabilities may allow automated applications that supersede the "power test" approach. Such prospects, however, raise a fundamental methodological misgiving.

Brief tests (and their emerging quantitative cousins) rely on holdings counts to both categorize the titles that comprise each test and then rank each library's collection. The OCLC database—conveniently, though unfortunately incorrectly—is assumed to represent any universe of relevant publications. To cite a specific case, 95 percent of about 1,300

newspaper and serial titles listed in one bibliography for provincial Peru are not represented in North American libraries. Can a repository in the United States measure its Peruvian collection against a sample drawn from and validated by a database that lacks so many materials? Evaluation methodologies constructed on such partial foundations will substantially misrepresent the universe of research resources.

A second area of doubt concerns the commensurability of the test methodology and its results as applied to distinct disciplines. Scientists and humanists, for instance, differ dramatically in their production and use of information. Library collections and services presumably vary too. Whether power tests can accommodate these differences has not been addressed.

The persistent library ideals of collections interdependence and collectively comprehensive coverage suggest another kind of reflection. More and more, we perceive library holdings as the pieces in a far-flung mosaic of sources held together by online bibliographic databases. Researchers can with increasing ease locate materials not held by their library. By now, the more general information available through the *Conspectus* may be irrelevant.

But what of librarians, as they continue to pursue collections cooperation? *Conspectus* proponents argue that assessments of existing collection strength are indispensable signals of current practice and future intent: "If you don't know where you are, you don't know where you are going, because where you are going is understandable only in relation to where you are." Zen metaphysics aside, libraries' purchasing power continues to shrink. The assumed yet unproven correlation between existing collections and current receipts may be weakening. The relevance of retrospective assessments to current cooperation may be waning as well.

More profound shifts are also under way. Scholarship is increasingly reliant

upon electronic materials that cannot be owned. Bibliographic records for cooperative resources such as the Center for Research Libraries' hardcopy collections are being added to local catalogs. As ownership becomes more difficult to define, local collections may no longer be particularly meaningful units for evaluation.

White's clear, provocative, and convincing account breaks new ground in a number of areas. But, as the author argues, collection evaluations are political artifacts as well as objective statements. In the final analysis, the politics of collections cooperation will determine whether the brief test methodology resuscitates the *Conspectus* as a collaborative tool. It's not at all clear that this should still be our goal.—*Dan Hazen, Harvard University, Cambridge, Massachusetts.*

The Reference Assessment Manual.

Comp. and ed. the Evaluation of Reference and Adult Services Committee, Management and Operation of Public Services Section, Reference and Adult Services Division (RASD) of ALA. Ann Arbor, Mich.: Pierian Pr., 1995. 372p. \$35 paper. (ISBN 0-87650-344x).

This publication is the product of a ten-year effort by a committee of the Reference and Adult Services Division (RASD) of ALA. It reflects a commitment on the part of a series of astute committee members and chairs to pull together a thorough inventory of instruments that have been developed in the field of reference services assessment. The thrust of this work is to support systematic and complementary assessment across libraries rather than leave the field to the current piecemeal approach. According to the preface, theirs is a two-part goal: (1) to provide one place for those in the field (practicing librarians, reference managers, researchers) to find all the instruments that have been developed for assessing reference activity, and (2) to encourage library administrators to support and promote evaluation of these services.

The structure and extensive scope of this manual should ensure that both goals are met. The fifteen chapters cover the full array of reference service components, for example, library users and reference patrons, reference environment, electronic databases and reference assistance, reference training, costs and outcomes, and reference effectiveness. Each chapter clearly defines its scope, conveys the importance of research in the area covered, and goes on to evaluate the state of the field; describes outstanding research needs; and lists instruments that are more fully described elsewhere. Also, each chapter's organization makes it easy to identify and explore particular areas of interest. For example, an administrator could easily pull out the information wanted on training or cost analysis, while a reference supervisor could work with the sections on duties and responsibilities or "question classification."

The Summaries of Instruments section is impressive. In most cases the actual instruments are provided on an accompanying disk (not available to this reviewer). When the instrument is not provided, full information about acquiring it is noted. The summaries also include information about reliability and validity testing for each instrument, when available (and specific experience with the instrument is also given). The clear intention is for new assessment to build on what exists. The work is also clearly intended to promote further testing of the reliability and validity of the instruments.

Another impressive section is the 140-page annotated bibliography that covers a broadly defined array of reference assessment works. Although it is presented as a "selected bibliography," it in fact offers extensive coverage of articles and monographs from the 1960s through the 1990s, as well as some unpublished works and, providing a historical perspective, some older items dating back as far as 1902. The sources cover both public and academic libraries and, more selectively,