On the Merits of Direct Observation of Periodical Usage: An Empirical Study

Marifran Bustion, John Eltinge, and John Harer

Measurement of current periodical usage is one of several tools used by librarians to make decisions regarding placement or deselection of serial titles. In an open-stack current periodical room, librarians may make such measurements through either direct observation of periodical usage or through less direct measures, e.g., reshelving counts or voluntary user responses. In principle, the direct-observation method is attractive because it may include broader classes of use than those covered by other measurement methods. Extended use of direct observation, however, is prohibitively expensive, and, if taken to an extreme, may itself intrude on total periodical use.

The authors conducted an empirical study to assess the practical merits of direct observation work. The resulting data indicate that observational errors, aggregation issues, and costs limit the feasibility of long-term, direct-observation use studies. With careful training of observers, however, direct-observation work may be useful in calibrating the results of other less expensive methods of measuring periodical use.

During the past few years, libraries have had to manage unprecedented increases in serials costs with limited or shrinking budgets. The effect on collection development has been a decrease in the number of monographs added, canceled serials subscriptions, increased interlibrary loan transactions, and decreased institutional research efforts.

In addition, cancellation of serials subscriptions invariably lead to some concerns among faculty and students. In some cases, specific departments or administrators may reasonably ask libraries to justify certain cancellation decisions. To provide such justification, and to ensure that cancellation decisions are indeed equitable and reasonable, it is useful to have available several objective measures of the

---

Marifran Bustion is Head, Acquisitions Department at the Gelman Library, George Washington University, Washington, DC 20052; John Eltinge is Assistant Professor, Department of Statistics; and John Harer is Head, Circulation Division at the Sterling C. Evans Library,Texas A&M University, College Station, Texas 77843-3143.

The authors express appreciation for the assistance of Irene Hoadley, Bill Hobson, Jane Oliver, Tommy Shaeffer, Richard Sellers, Tanya Wiggins, and Jo Ann Williams in the study. The authors also thank R. J. Bartlett and E. D. Alexander for computational work reported in the statistical analysis section.

This work was supported in part by grants from the Sterling C. Evans Library Research Committee and the Texas A&M University Mini-grant Committee. The work of the second author was also supported in part by grants from the National Institutes of Health (GM39015-01A1) and the A. C. Nielsen Company.

use of specific periodicals and groups of periodicals, and of the value of specific periodicals to a given institution.

Use studies are often designed to examine user questionnaires, interlibrary loan requests, photocopier use, reshelving counts, and direct and indirect observations of user behavior. All types of use studies have merit and rely to greater or lesser degrees on the cooperation or participation of users and library staff. User questionnaires rely heavily on user participation and cooperation; interlibrary loan and photocopier use depend on staff records; reshelving counts rely on user cooperation, but are more dependent on staff efficiency in record keeping and returning material to the shelves; and direct observation of user behavior relies heavily on staff involvement.

The authors designed the auxiliary study to assess the degree to which direct observation permits effective measurement of periodical use not recorded in reshelving counts.

Use studies are generally designed to address several questions, including estimation of total use and cost-per-use of specific periodicals; assessment of periodicals through faculty ratings and citation analyses; development of procedures for storage, cancellation, duplication, and acquisition of back volumes; and development of issue binding and retention schedules. Each of these issues plays a role in periodical cancellation decisions required by reductions in acquisitions budgets.[1] Questions involving shelf space or use density are of secondary interest.[2]

In the authors' use study, the principal measure of periodical use was a daily count of the number of reshelvings of each periodical. Some controversy exists regarding the use of a reshelving count as a proxy for the total use of a given periodical. For example, Colin R. Taylor reported a case in which reshelving counts represented only 22 to 40 percent of an alternative measure of journal use.[3]

The implications of a similar undercount rate for the Texas A&M University study are threefold. First, such an undercount may lead to a discouraging and misleading picture of the value of the library periodical collection to the university community. Second, the undercount may lead certain rarely used periodicals to be classified as unused. This misclassification might lead to removal of such rarely used periodicals under a deselection policy that employs zero use per se as an important criterion for cancellation.[4] Third, and most important, differences in the undercount rate across disciplines or periodicals may lead to reported reshelving counts that underrepresent or overrepresent the true proportionate use of specific periodicals or groups of periodicals. This result in turn may lead to inequitable or inefficient cancellation decisions.

These issues led the authors to consider direct observation of periodical use as an alternative to reshelving counts. In principle, a direct-observation study is more attractive than a reshelving-count study because the investigation may design a direct-observation study to measure specific use types of interest and avoid the differential undercount issues described above. More practically, however, observational errors, aggregation issues, intervention effects, and costs may limit the value of direct observation of periodical usage. To address these issues, the authors conducted an auxiliary study based on direct observation of current periodical usage in the Evans Library. This auxiliary study permitted the authors: (1) to assess the merits of the direct-observation method and (2) to compare the direct-observation use counts and reshelving counts for specific groups of journals on the same days. The present article addresses the first issue; subsequent papers will discuss the second.

The remainder of this paper is organized as follows. The Methodology section presents the basic methodology of the auxiliary study, including a description of the periodicals studied, an outline of the randomized design for the direct observation of stacks, and the definition of "use" employed in the direct-observation work. Methodology also discusses the assessment of observational errors and reports the costs of the direct-observation and reshelving-count studies.

The Statistical Results section reports some summary statistics and standard errors for the direct-observation method and also analyzes the structure of the observational errors encountered in this study. The same section discusses the degree to which these results may be generalized to other time periods and to titles and volumes not contained in the Current Periodicals Department. The section concludes with a discussion of some aggregation issues which arise in direct-observation studies. Finally, the Discussion section presents some general implications and limitations of the results of this study.

METHODOLOGY

The authors designed the auxiliary study to assess the degree to which direct observation permits effective measurement of periodical use not recorded in reshelving counts. The design involved observation of periodical use at randomly selected locations and times, a detailed operational definition of periodical use, and an assessment of the bias and variance of observational errors. Cost constraints were an important additional factor in the design of the auxiliary study.

The Scope of the Study

The investigators restricted the auxiliary study to the Current Periodicals Department (CPD) of the Evans Library. Recent issues of approximately 7,500 periodical titles were stored in the CPD. Throughout the study, the titles had an average age of approximately nine months (from date of receipt) and had an average of eight issues stored in the CPD. During the week of the study, approximately 1,600 issues were received and added to the CPD, and 660 volumes (approximately 2,640 issues) were removed for binding.

The auxiliary study did not consider titles or volumes outside the CPD. For one reason, differential undercount rates in reshelving measures were considered to be of particular concern for issues in the current periodical room. Different disciplines may make different uses of recently published periodicals. Some such uses, e.g., careful reading of articles, are more likely to be measured by a reshelving count than are other uses, e.g., casual browsing. In addition, uses like casual browsing are more prevalent in the Current Periodical Room than in the general stacks. Thus, direct observations of periodical use may be more appropriate for use in a current periodical room.

For another reason, the direct-observation method used in this study would be less efficient in the general stacks area, where periodicals and regular books are interspersed according to Library of Congress call numbers. For an auxiliary use study in the general stacks area, user-response methods may be preferable.[5]

Observational Design and the Direct-Observation Method

The auxiliary study recorded certain types of periodical use on randomly selected shelves at randomly selected times on three randomly selected days (Tuesday, April 17; Thursday, April 19; and Sunday, April 22) in the week of April 15–22, 1990. The direct observation method employed was similar to direct observation methods reported by Johanna Ross, Charles Wenger and Judith Childress.[6] The week chosen to be studied was not randomly selected, however. It occurred one month after the use study was half completed. Also, it was neither the first nor last week of a semester, but was at a time students were completing term papers and studying for exams. The authors considered it a typical week during the academic year.

Figure 1 presents the layout of the Evans Library CPD. In the discussion below, range refers to the side of a bookcase that contains periodicals, and aisle refers to the floor area adjacent to or between ranges of periodicals. Both the general and the auxiliary use studies excluded newspaper holdings. Thus, figure 1 displays a total of 51 relevant ranges arranged in 27 horizontal aisles.

		19B	RESHELVING AREA - A-QD		
		19A	QP/501/A18 - QR/360/J62		
9B	RESHELVING AREA - QE-Z	18B	QL/801/J9 - QP/474/V5	28B	Z/1037/A1/A5 - Z/8230.5/E44
9A	N/1/S9 - NX/765/A7	18A	QL/461/I5 - QL/801/A45	28A	Z/284/A9 - Z/1035/C5
8B	LB/1051/C678 - N/1/P83	17B	QK/1/S77 - QL/461/G49	27B	TS/1080/C34 - Z/278/S6
8A	K/25/N53 - LB/1051/B38	17A	QH/505/A1/H4 - QK/1/S68	27A	TP/757/P55 - TS/940/L4
7B	JA/4/C2 - K/25/N5	16B	QH/7/P495 - QH/505/A1/B8	26B	TN/677/A1/O412 - TP/700/S3
7A	HM/291/I595 - JA/3/P7	16A	QE/1/J8 - QH/5/B6	26A	TK/7870/E543 - TN/672/J68
6B	HG/11/M3 - HM/263/P767	15B	QD/271/J66 - QE/1/I7	25B	TJ/825/W5 - TK/7870/E54
6A	HG/11/B9 - HG/11/E2	15A	QD/1/A36 - QD/271/J65	25A	TD/201/W35 - TJ/810/S93
5B	HD/9000.1/S78 - HF/6/B43	14B	QC/461/J63 - QD/1/A35953	24B	TA/168/S9 - TD/201/W345
5A	HC/186/A1 - HD/9000.1/J6	14A	QC/1/P45 - QC/451/I5	24A	T/1/A66 - TA/168/I5
4B	HB/1/E56 - HC/167/S67/L37	13B	QA/801/J682 - QC/1/P43	23B	SF/208/D37 - SK/601/A1/C36
4A	GV/975/G7 - HB/1/E525	13A	QA/76.5/N8 - QA/801/A7	23A	SB/599/I4 - SF/201/A6
3B	GB/841/J6 - GV/975/G6	12B	QA/1/J45 - QA/76.5/M522	22B	S/604.8/R39 - SB/599/F52
3A	E/185.5/J8 - GB/841/I2	12A	Q/11/I5 - QA/1/I92	22A	RK/71/J6 - S/601/A37
2B	D/839/C87 - E/184/S75/H5	11B	PR/9100/C25 - Q/11/F65	21B	R/11/B7 - RK/1/B2
2A	BF/233/P4 - D/839/C837	11A	PN/2091/E4 - PR/8700/J68	21A	New York Times - Die Zeit
1B	AP/63/H56 - BF/204.5/J68	10B	PF/3001/S4 - PN/2081/R4/L5	20B	El Paso Herald - New York Times
1A	AG/305/A4 - AP/63/C6697	10A	P/1/A1/C22 - PF/3001/G3	20A	AFL-CIO News - El Diablo

FIGURE 1
Layout of the Evans Library Current Periodicals Department

Sunday, April 22, 12:00 - 12:05 Observer:________________

5B --Inside: ________________

5B --Outside: ________________

6A --Inside: ________________

6A --Outside: ________________

Sunday, April 22, 12:05 - 12:10 Observer:________________

11B --Inside: ________________

11B --Outside: ________________

12A --Inside: ________________

12A --Outside: ________________

FIGURE 2
Time and Aisle Assignments

Twenty-four of the aisles adjoin 2 ranges (e.g., 7B and 8A), while the remaining 3 aisles adjoin a single range (e.g., 1A).

Since time of day is an important factor in periodical use, the investigators partitioned the CPD operating hours into the following blocks: Tuesday, April 17 and Thursday, April 19 (8:00 a.m. to midnight), 6 periods of 160 minutes each; Sunday, April 22 (noon to midnight), 4 periods of 145 minutes each, and 1 period of 140 minutes. Within each time block, each aisle was assigned at random to 1 or 2 five-minute segments. The columns of figure 2 reproduce an example of the resulting time segments and aisle assignments. The statistical literature describes this type of stratified sample design as controlled selection.[7]

Student workers observed the selected aisles in the indicated time segments. Preliminary work indicated that an observer could distinguish clearly between use in the right and left ranges of an aisle and between use of periodicals in the inside and outside of each range. (The rules define the inside and outside of a range to be the halves of ranges closest to specific walls in the room.) The rules, listed below, explained to observers how to identify a use and record it:

1. Enter name in space marked Observer.
2. Make a single hash mark, /, each time you observe one of the types of count described below.
3. Be sure to mark your counts on your data form separately for the inside and outside half of each range, as indicated on the form. "Inside" is the end of the range closest to the wall by Acquisitions; "outside" is the end of the range closest to the window. Unless told otherwise, please observe from the outside end of the range.
4. Count any periodical issues that are returned to the shelf by any person who does not work in CPD. If a single person returns more than one issue to the shelf, write down your best estimate of the number of

issues which that person returned to the stacks.

5. Count each time a person uses a periodical, even if he or she just picks up an issue and looks at the cover.
6. Count as one use a single person's use of one or more issues from a single stack of issues. If the patron removes an issue after picking up a stack, count as one use each issue removed and do not count the issues returned. If the patron does not remove any issues from a stack, but returns the entire stack to the shelf, count that stack as one use.
7. Count separately each time the same person uses periodicals from separate stacks of issues.
8. Count only the uses completed during the five-minute period that you observe a particular stack.
9. Do not count periodicals a person carries away from the aisle you are watching.

Following the rules reproduced above and discussed in the Definitions section, the observers recorded separately the number of uses in the inside and outside halves of the right and left ranges of the observed aisle. These observations were recorded in the indicated columns of figure 2. To reduce the effect of observer fatigue on recorded use counts, no student worker observed aisles for more than two hours at a time.

Definitions of "Use" and Observational Rules

In defining use of periodicals for the direct-observation method, the authors assumed that patrons reviewed issues in the aisles to locate an issue containing a specific article and to locate information in a specific article or as a result of a random search (browsing). A casual review or browsing of issues may yield some useful information. Since observational techniques are generally unable to distinguish actual value of each review, each casual review was considered important and was counted. For the purposes of this study, casual use was defined as any look at the cover, contents page, or information contained in the body of the issue. A glance at the spine for volume number or date or handling of any issue to retrieve another issue was not considered a use.

To implement the direct-observation method, the investigators developed the direct-observation counting rules reproduced above. The focus of these rules was to have an observer count every use of periodicals in an observed aisle, except for uses which subsequently would be recorded in a reshelving count. Some distinctions were required from the operational definition of use outlined in the rules. For example, rule six gives a single count to one patron's use of one or more issues stored in a single stack. Otherwise, a patron's brief perusal of all of the issues in one stack would lead to an inflated count of use. Also, rule four directs the observer to estimate and record the number of issues returned to an observed aisle by a patron. The investigators reasoned that patron removal and return of several issues of the same title indicated a greater intensity of use than the casual scanning of several issues considered in rule six. Moreover, if the patron had not reshelved the issues, each of the issues would have been included in the reshelving count. Thus, rule four is intended to parallel the implicit counting rule employed in the reshelving count. Finally, rules eight and nine address the limited observation times allocated to each aisle. Some preliminary observations indicated that some uses of a single issue in the aisles lasted one minute or more. In the absence of rule eight, the random observational design would give such long uses a higher probability of being recorded than the short uses. To parallel the reshelving counts, the investigators chose to use rule eight to ensure that all patron uses of issues have the same probability of being recorded, regardless of the length of use. Similar reasoning applies to rule nine.

Assessment of Observational Errors

In assessing the merits of a direct-observation study of periodical use, it is important to measure the magnitude of observational errors associated with misinterpretation of rules by the student workers. To address this issue, the three investigators observed the same aisles as student workers in some randomly selected five-minute segments. Student workers recorded each observed use with a single hash mark (/), and the faculty members recorded each use with one of several symbols: S (scan), C (careful use), or R (reshelved by patron). A fourth use, T (taken away by patron), was also recorded by the investigators but is not directly relevant to the student workers' observations. Note that observational rule nine for student workers specifically excludes the T usage recorded by faculty members. Separate recording of these four use types allowed the investigators to study whether observational errors were associated primarily with a particular use type. For example, an investigator might speculate that careful use (C) is more clearly defined, and thus less subject to observational error than scanning (S). The Statistical Results section provides an empirical discussion of this issue.

For the remainder of this paper, the sum of the faculty S, C, and R counts for a given half-range will be defined as the true count of use (excluding staff reshelving of issues). The difference between the student use count for a given half-range and the corresponding true count equals the observational error in the student count. From a statistical point of view, this definition is restrictive, because it excludes the possibility that the investigators' observations may also contain measurement error. Nonetheless, this approach appears to be reasonable. The faculty members had a vested interest in the project; one generally associates such interest with greater attention to detail. In addition, the faculty observers were more familiar with the observational rules and the study's purpose. Consequently, the investigators expected fewer errors in faculty observations than in student observations. Use of a more elaborate errors-in-variables model, which would allow for both student and faculty observational errors, is beyond the scope of the present work.[8]

Costs of the Direct-Observation and Reshelving-Count Studies

The direct-observation study used a total of 53 student-worker hours. At a pay rate of $3.80 per hour, the direct cost in student wages was $197.60. In addition, the authors spent 20.5 hours in direct observation. If a direct-observation study were conducted using one student worker-hour for each of the 111 CPD operating hours per week, the cost, at $3.80 per hour, would be $421.80 per week.

During the same week, the CPD used 50 student hours to record reshelving counts. These 50 hours were in addition to the 22 hours per week otherwise required to reshelve periodicals in the CPD. Using the same pay rate, $3.80 per hour, the marginal student worker cost for recording reshelving counts was $190, a daily cost of $27.14.

STATISTICAL RESULTS

Assessment of the statistical results of this study requires considering basic descriptive statistics for the observed use counts; the distribution, bias, and variance of errors in the observed-use counts; limits on the generalizability of the reported results; and aggregation issues. A paper by Bartlett and Eltinge presents a more detailed statistical analysis of the data from this auxiliary use study.[9] In that paper Bartlett and Eltinge discuss some regression and variance component models for the faculty and student worker observations; consider alternative models based on logarithmic data transformations and trim- ming of extreme observations; and evaluate least-squares methods of predicting true usage counts based only on student worker observations.

Counts by Student Workers and by Faculty

Under the random design outlined above, student workers observed a total of 615 aisle-time combinations. Observation of a given aisle led to use counts for two or four half-ranges in the specified time segments. For 792 of these half-range student observations, there was a matching faculty observation for the same half-

TABLE 1
FREQUENCY DISTRIBUTION OF FACULTY AND STUDENT COUNTS

Observed count	0	1	2	3	4	5	6	7	8	9	10 or more
Faculty frequency	734	21	7	5	0	6	2	4	0	5	8
Student frequency	739	17	6	3	5	1	4	5	2	0	10

TABLE 2
FREQUENCY DISTRIBUTION OF THE ERRORS, (STUDENT COUNT)– (FACULTY COUNT) FOR THE 58 NONZERO FACULTY OBSERVATIONS

Observed error	–9 or less	–8	–7	–6	–5	–4	–3	–2	–1	0
Frequency	4	1	2	1	4	2	2	5	13	8
Observed error	1	2	3	4	5	6	7	8	9 or more	
Frequency	6	0	2	3	1	1	1	0	2	

range at the same time. Table 1 presents the frequency distributions of the student counts and faculty counts for these matched observations. Note that for both student and faculty observations, a large majority of the use counts equalled zero. Table 1 also indicates substantial differences between the frequencies of student and faculty counts. As noted in the Methodology section, the authors used the working hypothesis that faculty observations were the true use counts. Thus, observational error is defined as any difference between a student count and a corresponding faculty count. The authors expected the observational error process to be fundamentally different for zero and nonzero true counts. For example, with a true count equal to zero, a student observation cannot be less than the true count. Consequently, the authors discuss below the differences between student and faculty observations for the 58 paired observations with a nonzero faculty observation; some possible explanations of variability in the same set of observational errors; student observations for the 734 paired observations in which the faculty observer recorded no auxiliary periodical use; and other statistical aspects of this study.

Observational Errors for Nonzero Faculty Counts

Table 2 reports the frequency distribution of the differences, (student observation) - (faculty observation) for the 58 matched observations with a nonzero faculty count. For these 58 pairs, the faculty counts had a sample mean of 5.41 and a sample standard deviation of 6.15; the student counts had a sample mean of 4.39 and a sample standard deviation of 8.01. The differences between the student and faculty counts had a sample mean of -1.02 and a sample standard deviation of 5.29. A formal test of the null hypothesis of no difference between the overall mean student count and the overall mean faculty count, against the alternative hypothesis of a nonzero difference between these two means, led to a t-statistic equal to -1.47 on 57 degrees of freedom. This test statistic was not significant at the 0.10 level of significance. In practical terms, this means that sufficient evidence does not exist to conclude that the student observations were systematically higher or lower than the corresponding faculty observations.

Given this conclusion, we may study the variance of the differences between the matched student and faculty counts. One useful measure of this variability is an estimated reliability ratio, defined in this case as equaling the sample variance of the true observations, divided by the sample variance of the student observations.[10] For the 58 matched pairs under consideration, this estimated reliability ratio equals 0.59, with an estimated standard error equal to 0.19. In practical terms, this point estimate of the reliability ratio suggests that about 59% of the variability of the student counts is attributable to variability in the true counts, and the remaining 41% of variability in student counts is attributable to measurement error. However, the large standard error indicates the relatively poor precision of this point estimate. The estimated reliability ratio of 0.59 is not entirely encouraging, but also is not entirely out of line with reliability ratios for some social-science measurements based on complex concepts.[11]

Relationships between Observational Errors and True Use

To study further the observational error question, figure 3 presents a plot of the 58 student observations against the corresponding nonzero faculty observations. Strike-overs in the lower left section of the plot indicate multiple observations at the same points. The plotting symbols are letters representing the different student observers. If there were no measurement errors, the matched student and faculty observations would be equal, so that all plotting symbols would fall on a straight line with a slope of one and an intercept of zero. The distance of a given point above or below this straight line indicates the magnitude of measurement error in a given student observation.

Define Y_{ijk} to be the use count recorded by the *ith* student for the *kth* half-range in the *jth* aisle, and define x_{ijk} to be the corresponding faculty count. Then a linear regression model (see chapter 1 of the textbook *Applied Regression Analysis* by N. R. Draper and H. Smith,[12]) for the Y_{ijk} and x_{ijk} observations is,

$$Y_{ijk} = b_0 + b_1 x_{ijk} + e_{ijk}$$

MODEL 1

where b_0 and b_1 are the fixed intercept and slope of a straight-line model relating Y_{ijk} to x_{ijk}, and e_{ijk} is a residual term accounting for random variation in the student observational errors. The conjecture that there was no systematic undercount or overcount in the student observations is equivalent to the null hypothesis that $b_0 = 0$ and $b_1 = 1$. An F test of this null hypothesis against the alternative hypothesis that (b_0, b_1) does not equal (0,1) had a test statistic $F = 1.07$ on 2 and 56 degrees of freedom, and was not significant at the 0.10 level of significance. An estimated generalized least squares fit of model 1 with a variance component model for random student, aisle, and half-range effects led to an estimate of $b_1 = 0.98$, with a standard error of 0.10; and an estimate of $b_0 = -1.11$, with a standard error of 1.63.[13]

In addition, note that in figure 3 observations by certain students appear to be systematically higher (e.g., student A) or lower (e.g., student B) than the corresponding faculty observation. However, for a variance component model for the residuals of the regression model described above, a formal analysis of variance test of a "student effect" led to an F test statistic equal to 1.96 on 9 and 7 degrees of freedom. This test was not significant at the 0.10 level of significance.

As noted in the Methodology section, the true use count is the sum of observed S (scan), C (careful use), and R (reshelved by patron) counts. Review of the counting rules listed above suggests that errors in the student observations differ across different types of use. Estimation of two regression relationships helps evaluate this suggestion. First, an ordinary least squares regression of the errors $Y_{ijk} - x_{ijk}$ on the corresponding faculty S_{ijk}, C_{ijk}, and R_{ijk} counts led to the estimated equation,

$$Y_{ijk} - x_{ijk} = -1.00 - 0.24\, S_{ijk} + 1.10\, C_{ijk} - 0.44\, R_{ijk} + d_{ijk}$$

MODEL 2

where the error term d_{ijk} will be discussed further below. The intercept and coefficients for S_{ijk}, C_{ijk}, and R_{ijk} had estimated standard errors equal to 0.88, 0.12, 0.33, and 0.31, respectively. An informal interpretation of model 2 is that, after accounting for the effects of the other variables in the equation, and after

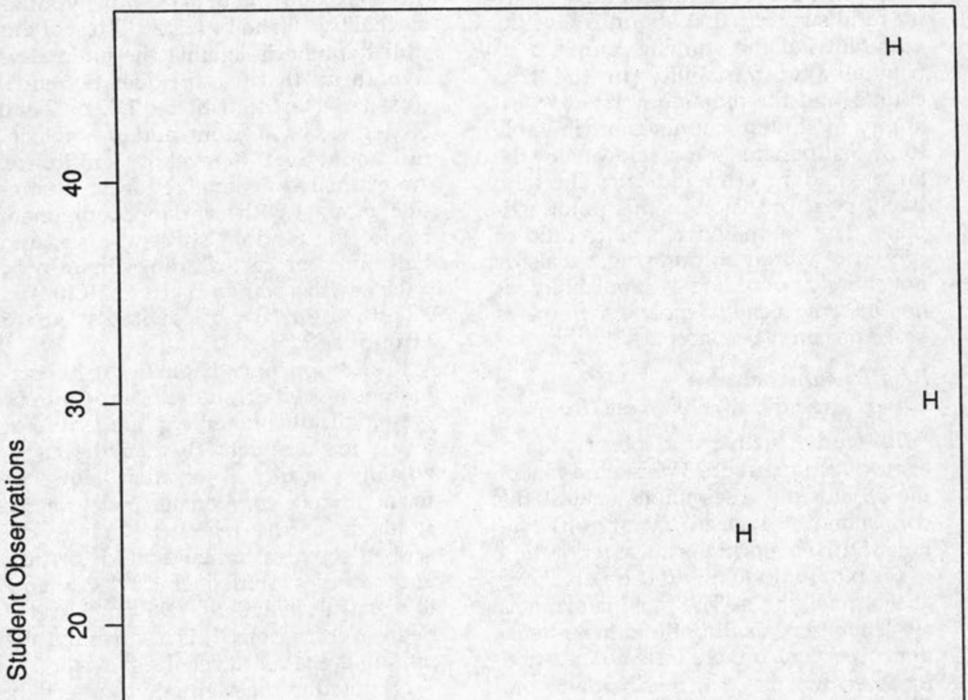

FIGURE 3
Plot of Student Observations against the Corresponding Nonzero Faculty Observations

TABLE 3
FREQUENCY DISTRIBUTION OF STUDENT OBSERVATIONS WHEN THE CORRESPONDING FACULTY OBSERVATION IS ZERO

Student observation	0	1	2	3	4	5	6	7	8
Frequency	719	4	2	1	3	0	2	2	1

accounting for an overall undercount or overcount rate, a significantly positive coefficient for a given independent variable (C_{ijk}, at the 0.01 level) indicates that the student workers tended to overcount this type of use, while a negative coefficient (for S_{ijk}, at the 0.05 level) indicates that student workers tended to undercount the corresponding type of use.

By contrast, if there were no differentially systematic undercount or overcount of any use type, we would expect each of the slope coefficients in model 2 to equal zero. A test of the null hypothesis that the S, C, and R coefficients were all equal to zero had an *F* test statistic equal to 11.24 on 3 and 54 degrees of freedom; this test was significant at the 0.01 level of significance.

Second, we may expect the variability in observational errors to vary across types of observed use. To address this issue, the squares of the residuals d_{ijk} from model 2 were regressed on the faculty S_{ijk}, C_{ijk}, and R_{ijk} counts. The resulting estimated model was,

$$d_{ijk}2 = -0.06 + 4.18\, S_{ijk} + 6.95 C_{ijk} - 0.09\, R_{ijk} + \text{error}$$

MODEL 3

with the estimated standard errors for the intercept, and the coefficients of S_{ijk}, C_{ijk}, and R_{ijk} equal to 6.2, 0.89, 2.3, and 2.2, respectively. An informal interpretation of model 3 is that independent variables with coefficients significantly greater than zero (S_{ijk} and C_{ijk}, each at the 0.05 level of significance) are the use types associated with greater variability in nonsystematic observational errors.

In addition, if variability in nonsystematic observational errors was the same for each use type, then we would expect each of the slope coefficients in model 3 to equal zero. A test of the null hypothesis that the S, C, and R coefficients were all equal to zero in model 3 had an *F* test statistic equal to 13.94 on 3 and 50 degrees of freedom. This test was significant at the 0.01 level of significance.

Observational Errors for Zero Faculty Counts

Table 3 reports the frequency distribution of the student observations matched with faculty observations that equalled zero. Note that for 719 of the 734 faculty zero counts, the student also recorded a zero count. The relatively low frequency of student overcounts in this case is not surprising because, for most faculty counts equal to zero, there were no patrons in the aisle under observation.

For the remaining 15 cases, the student recorded a use count between 1 and 8. It appears that most of these overcounts were associated with misinterpretation of rules four, six, and nine for times in which a patron was in the observed aisle.

Generalization of Results

This study, like many use studies, does not permit one to make formal statistical inferences to populations other than the CPD and period studied. Generalization of the results reported here depends on the comparability of the Evans Library to other libraries in terms of definitions and measures of periodical use, training of direct-observation workers, and periodical use patterns.

First, variability in administrative interests may lead to legitimate differences in definitions and measures of total periodical usage. For example, as noted in the Methodology section, the definition of total use of current periodicals depends substantially on the degree to which one recognizes casual use. Second, the specific statistical results on observational errors apply only to observers with general backgrounds and observational-rules training similar to that of the student observers. Since the student workers employed in the direct-observation study were regularly assigned to other duties in the Evans Library CPD, they had some familiarity with library patrons and periodical use. Their training in the direct-observation rules, however, involved only a brief (thirty minutes) review of the observation rules. Given the relatively small reliability ratio reported in the Statistical Results section, it is clear that more extensive training for workers in a long-term direct-observation study is preferable. A separate reliability study could then assess error magnitudes for these more extensively trained observers.

Third, periodical use patterns may themselves influence the extent to which observational errors are of practical concern in a direct-observation use study. As noted in the Methodology section, observation of a single careful use is fairly straightforward and is generally unlikely to result in a measurement error. Casual use, however, requires the observer to exercise some judgment and thus is more likely to result in measurement errors. Thus, an investigator is more likely to be concerned about observational errors for libraries and times in which casual use is a substantial component of periodical use.

Aggregation Issues

Aggregation places an additional constraint on the utility of a direct-observation study. Some measures, such as reshelving counts or voluntary user responses, associate a use count with a specific periodical title or issue. Other measures of value, such as citation analyses or faculty ratings, are also title-specific. Thus, such measures may be used fairly directly in a journal placement or deselection decision. As noted in the introduction, however, reshelving counts, user responses, or other nondirect measures may not reflect several important forms of periodical use.

By contrast, an investigator may define direct-observation rules to include browsing or other forms of casual use, but direct-observation counts are recorded only for total use within a given half-range in a given time period. Observations recorded for a finer level of aggregation, such as for a quarter-range or for individual titles, do not appear to be feasible. Since each half-range includes sixty-five to eighty titles with consecutive Library of Congress call numbers, a single direct-observation count generally measures aggregate use within one or more related subdisciplines, but gives no specific information about the use of specific journals.

Thus, direct-observation use counts are most likely to contribute to journal acquisition, placement, and deselection decisions by indicating the relative journal use intensities in different subdisciplines and by permitting librarians to assess the differential undercount issue raised in the introduction. A detailed discussion of these two is beyond the scope of the present work. Another paper will present some specific data analyses for these issues.

DISCUSSION

Training and Attentiveness of Observers

The results of the auxiliary use study indicate that student observations of periodical use contain a substantial component of observational error. The question arises as to how such errors could have occurred. The first and most obvious hypothesis is that the training was not adequate. As noted, the training session was brief, about one-half hour at most. During the training sessions, most observers appeared to have understood the rules of the project. The majority of the student observers were quite familiar with library practices and procedures. Many were shelvers in the CPD. Nonetheless, the short training period and the training method may not have been sufficient for the absorption of a complex set of rules.

Second, boredom may have contributed to measurement error problems. As indicated in table 1, the direct-observation method required an observer to spend substantial amounts of time recording zero use. Naturally an observers' attention wandered during such periods. For the short periods in which periodical use did occur, the observer needed to record use counts carefully, according to some fairly complex rules. This combination of inactivity and need for careful attention to detail may have contributed substantially to observational error problems.

The observers attempted to be unobtrusive in as many instances as possible while still holding some confidence that the patron could be observed accurately.

Intervention Effects

When the study was designed, there was some concern that the direct-observation method would have some deleterious effect on the patrons using the CPD. If so, such an effect would not only bias the results of the study but, more importantly, would destroy the ability of the CPD to deliver its service effectively. If the intervention made patrons self-conscious to the point that they chose not to search the CPD shelves at all, then direct observation would have created a negative effect that would not have been present otherwise. The observers attempted to be unobtrusive in as many instances as possible while still holding some confidence that the patron could be observed accurately. While this was not always possible, the investigators did not note any evidence to suggest a negative intervention effect on patron use of periodicals. For example, the investigators did not note any cases of patrons staring at the direct-observation staff or moving out of an aisle when a patron saw an observer.

CONCLUSION

The direct-observation method earned mixed reviews. Direct observation of periodical use is attractive because observational rules may be tailored to satisfy specific administrative definitions of use. However, observational errors, aggregation issues, and costs may limit the administrative utility of direct-observation use counts. Librarians may reduce the observational errors problem through additional training of observers. Aggregation and cost issues, however, are fundamental constraints on the value of the direct-observation method.

For the Evans Library use study, the authors now plan to use direct-observation counts primarily to assess differential undercount rates for other use measures, such as reshelving counts. Another article will present details of this assessment.

REFERENCES AND NOTES

1. Studies addressing concerns of use studies include Robert N. Broadus, "The Measurement of Periodicals Use," *Serials Review* 11:57–61 (Summer 1985); Maurice B. Line and Alexander Sandison, "Practical Interpretation of Citation and Library Use Studies," *College & Research Libraries* 36:393–96 (Sept. 1975); Allen Kent et al. *Use of Library Materials: The University of Pittsburgh Study* (New York: Dekker, 1979); Katherine Konopasek and Nancy Patricia O'Brien, "Undergraduate Periodicals Usage: A Model of Measurement," *Serials Librarian* 9:65–74 (Winter 1984); and Marifran Bustion and Jane Treadwell, "Reported Relative Value of Journals versus Use: A Comparison," *College & Research Libraries* 51:142–51 (Mar. 1990).
2. Studies addressing shelf space and use density questions are Dianne C. Langlois and Jeanne V. Von Schulz, "Journal Usage Survey: Method and Application," *Special Libraries* 64:239–44 (May/June 1973); Charles B. Wenger and Judith Childress, "Journal Evaluation in a Large Research Library," *Journal of the American Society for Information Science* 28:293–99 (Sept. 1977); and Line and Sandison, "Practical Interpretation," p.394.
3. Colin R. Taylor, "A Practical Solution to Weeding University Library Periodicals Collections," *Collection Management* 1:27–45 (Fall/Winter 1976–77).
4. Ibid, p.34–41.
5. Langlois and Von Schulz, "Journal Usage," p.240; Taylor, "A Practical Solution," p.31–32.

6. Johanna Ross, "Research Notes: Observations of Browsing Behavior in an Academic Library," *College & Research Libraries* 44:269–76 (July 1983); Wenger and Childress, "Journal Evaluation," p.294–95.
7. For a general discussion of controlled selection designs, see W. G. Cochran, *Sampling Techniques* (New York: Wiley, 1977), p.124–27; and J. J. Waterton, "An Exercise in Controlled Selection," *Applied Statistics* 32:150–64 (1983).
8. For a discussion of errors in variable models, see W. A. Fuller, *Measurement Error Models* (New York: Wiley, 1987).
9. R. J. Bartlett and J. L. Eltinge, "Variance Component Approaches to Measurement Errors in Count Data: An Illustrative Example," unpublished manuscript, Department of Statistics, Texas A&M University.
10. For a more detailed discussion of the reliability ratio and related issues, see Fuller, *Measurement Error Models*, Section 1.1.2.
11. For example see table 1.1.1 in Fuller, *Measurement Error Models*, p.8.
12. N. R. Draper and H. Smith, *Applied Regression Analysis*, (New York: Wiley, 1981), Chapter 1.
13. Generalized least squares fit discussed in Draper and Smith, *Applied Regression Analysis*, Section 2.11.