USE OF A DIARY AND REPLICATED SAMPLING TO ESTIMATE FACULTY WORKLOAD

Howard B. Christensen, Bruce A. Chadwick, David B. Magleby,
Howard B. Christensen, Brigham Young University, Provo, UT 84602

Key Words: Diary, Replicated Sampling

Introduction
The genesis of this study, like so many, had its roots in a study of state institutions whose mandate was "to develop a faculty profile at the various institutions in the system of higher education regarding a typical faculty workweek." (McCarrey, 1978). Such a study was conducted in 1978 and again in 1990 among public institutions in the State of Utah. A desire for comparative data from a large private institution is partly responsible for the undertaking of our study.

Studies of faculty workload are not new and have been conducted over many years and in different locations. In 1990, The National Center for Education Statistics published their report of a nationwide survey of higher education entitled, *Faculty in Higher Education Institutions, 1988*. The design for this study involved a stratified, multi-stage sample with strata defined by level of degree offered and the size of the institution. Faculty from a given institution were selected using a multi-stage sampling procedure considering full and part-time faculty along with program areas. The sample was selected to provide unbiased estimates and the study had an overall response rate of 76 percent. More specific details can be found in Appendix A: Technical Notes of the publication, *Faculty in Higher Education Institutions, 1988*.

In the latest Utah study, concern about the quality of the study was expressed by some of the study participants because each institution was allowed to collect its own data with no clear protocol established for collecting the data. In addition, the instrument used in that study and in most other studies relied on recall and estimation of time spent on various tasks. For instance, the questionnaire used to collect the data on workload asked the respondent to report "on the average, how many hours per week did you spend at each of the following kinds of work." Or, "Please estimate the percentage of total working hours", etc. Therefore, memory or lack of it would provide only estimates of workload. Though the national study (Center for Education Statistics, 1990) used a much better design and exercised more control over the method of data collection, it still relied completely on recall of activities and workload.

We wanted to minimize memory recall problems and to provide a clear method of reporting and collecting the data. The following report reflects our successes and failures in this regard.

Methodology
We anticipated at the outset of the study that there would be resistance from some faculty members who felt we were intruding on their time and that they were being asked to justify their employment. To allay those fears as much as possible, we sought some initial feedback from the faculty to find out what types of approaches would elicit the most cooperation. We organized a series of about 10 focus groups from faculty on campus during the summer prior to the semester in which the study was conducted. In forming the focus groups, we, insofar as possible, randomly picked participants and tried to make sure that we covered the spectrum of colleges at the university.

This was very illuminating for us. It revealed a wide variety of attitudes along with intensities of feelings. Some faculty members strongly objected to the whole concept of measuring work and productivity. They felt the data would not be reflective of what really goes on; that there was no way to accurately measure the things we wanted to measure. Others viewed it as an opportunity to validate their work. Others felt it would be a good exercise to help them personally see how they spent their time and thus help them become more efficient in the use of their time from day to day.

Before we met the focus groups, we anticipated that some kind of daily diary or time-log would be the most logical approach to collect workload information. Therefore, we sought reactions to that approach as we explained our purpose to each group. Again, we experienced different reactions. Some objected to the extra time it would take to fill it out. (We suggested that it would need to be filled out each day, at least, to make it as accurate as possible.) Others suggested that we assign a student "tag" to them through the day who would follow them around each day with the purpose of recording their activities. This suggestion was not uniformly received and had obvious problems. There would be tasks the "tag" could easily recognize. But there would be other tasks that would be hard to classify by anyone other than the faculty member.

Another suggestion was to equip persons with a beeper and at certain times through the day, the faculty member would be beeped and asked to respond with the activity on which the person was engaged.

After all of these ideas were examined, we still felt that the daily diary would be the simplest and most easily implemented method for collecting the data. And that, regardless of the method we chose, there would be those who would object and probably refuse to cooperate.

However, judging from the reaction of participants in our focus groups, we didn't think we could ask participants in the study to keep a log for more than a week. Therefore, the log was set up to cover the days from Monday through Saturday. (Sunday was omitted because of the nature of the university and its expectations that Sundays would and should be used for activities that would not necessarily be university related.)

We anticipated that some weeks in the semester might be busier than others. Therefore, to get an accurate reflection of the "average" work week we decided to collect data for each of the 16 weeks of the semester. Thus, we would select a separate sample of respondents (panel) for each of the 16 weeks. Since, we did not expect cooperation beyond a week's time, each panel in the sample was independent of the other panels.

With this approach, we could obtain an independent estimate of total time spent for each of the 16 weeks, which could be summed to provide an estimate of the entire semester, or averaged to get an "average-week" estimate.

Developing the Instrument

During the time we were covering issues in the focus groups about the method of collecting workload data, we were likewise concerned about the types of classifications of tasks to include in the log. Teaching, research, and administrative activities were obvious. But, other classifications were also included: professional presentations, professional committees and university government, community service, and an area that would probably be unique to the university being studied, church service. An "other" category was included to cover options we had been unable to anticipate. A list was included with the time log for each of these broad classifications that helped the respondent more clearly classify his/her time. For instance, under the teaching category were included such items as:

Classroom instruction
Grading independent study courses
Thesis advising, reading senior papers, honors theses, etc.
Office hours for student advising and mentoring
Meeting with teaching assistants
Course preparation
Curriculum development
Unscheduled teaching (guest lectures, thesis committees)
Writing letters of recommendation
Administering examinations, both written and oral
Grading exams and papers
Reading and collecting reference materials for course preparation

Conferences and workshops on topics from professional expertise
Honors classes and related tasks
Supervising student teachers and interns

A total of 38 different activities were listed under the main classifications, covering as many university activities as possible. Input from the members participating in the focus groups was invaluable in identifying these sub-classes. For instance, we found the tasks of the fine arts faculty to be quite different from those in the more traditional, math, chemistry, and English areas of the university.

Along with the time log, we also prepared another instrument which we called a Time and Activity survey. This instrument was more similar to the instrument used in the national study previously referred to and relied upon recall since we felt the issues covered in this instrument were less likely to be hampered by memory problems. For instance, this instrument was to collect productivity measures such as number of courses taught during the preceding Fall and Winter semesters, the number of theses directed, number of books authored, papers published, etc. Sixty-eight variables were identified to be included in this instrument with the focus groups providing suggestions as to what should be included to cover the broad spectrum of activities encompassed by faculty across the entire university. We felt the faculty would have a very good grasp of these measures and could give, in turn, an accurate response.

This instrument collected activity and productivity measures from the point of view of "number of items" produced as well as "average number of hours spent in a typical week (month, semesters, etc.)" to measure the faculty member's recollection of what they accomplished beyond the measure of hours-on-task that would be collected by the daily log.

The major categories included in this instrument included: (1) Teaching, (2) Research and creative works, (3) Professional service and presentations, (4) Professional committee work and university government, (5) Community service, (6) Church service, (7) Administration, and (8) Other. Each of these categories was subdivided, as suggested above, into two categories: Activities and Time. Under activities, we collected "how many" activities of various sorts were accomplished and under time we collected "how much time on average" was spent on various tasks.

The Sampling Scheme

The sample design for the Faculty Time survey might best be described as a rotating panel design with no overlap from panel to panel. The sample within a time period (one week of a 16-week semester) was a stratified random sample with the strata being based on college and time-in-service at the university. Colleges

were grouped into 8 groupings of approximately equal size, and time-in-service was subdivided into three groups of 0-5 years, 6-15 years, and 16+ years. Thus, each faculty member would belong to one of a total of 24 strata determined by college and time-in-service.

A random sample of 20 faculty was selected from each of the 24 strata producing a total sample of 480. These groups were essentially randomized and a member of each group was assigned to panel number 1 through panel 20. The first 16 panels were of interest since each panel would be asked to keep a diary of activities for one of the weeks of the 16-week semester. Members of panels 17-20 were used as replacements for those individuals that were identified as "problems" in any of the initial 16 panels. Such problems were faculty who were, for instance, on professional-development leave during the semester to be studied, or had major assignments other than their university assignment during the semester. Overall, we made very few replacements because of such problems.

Every reasonable effort was made to enlist the cooperation of faculty members chosen to be part of the study. A letter under the signature of the president of the university was sent to each faculty member selected to participate. Reminders were sent to each participant after they received their copy of the weekly log. But as usual, there was a problem of non-response. If a person was assigned to keep a log for a week and then failed to do it—even though they were willing to keep the log for another week—we decided to excuse them from participating.

You recall that there were really two instruments used in this study: the daily log which spanned a week's time and the productivity instrument. Though we had selected the entire sample before the beginning of the semester, our general strategy was to contact the sample of respondents during the latter part of the week prior to the week for which they would be keeping their daily log. However, the productivity instrument didn't have to wait for a particular week in order for the respondent to answer it. Therefore, after the press of the first part of the semester was past, the productivity instrument was distributed to all participants. Those who had already filled out the daily log had been warned that an additional instrument would be coming. Those who had not were informed that they had been selected to participate in this entire study and that at some future time in the semester, they would receive a daily log and would be asked to keep a record of their time spent on university-related tasks for a week's duration.

Fieldwork Supervision
The day-to-day operation of data collection was carried out by a small staff of experienced persons. A packet of materials was prepared for each new panel of respondents and mailed to them so they received them on the Thursday or Friday of the week prior to their

participation. As previously indicated, each packet enclosed a letter under the signatures of the president of the university and an academic vice-president encouraging their participation in the study. An envelope for returning the completed study was included with the instrument, and all mailing was handled through the campus mail delivery system.

Participants whose instruments were not received were called and asked to return them as soon as possible. In some instances, these individuals indicated that they did not keep a log and promised to do it for the next week. To maintain the integrity of the individual weekly samples, we kindly thanked them for their offer but told them that we could not use their information. Usually only one or two refused to participate. Others had simply put their completed log aside and sent it in after the reminder.

Weekly Response Rate
In each of the 16 weeks of the study, 24 respondents were selected for an overall sample size of 384. As would be expected, some individuals did not return their questionnaires. With respect to the daily log, we couldn't use a typical follow-up procedure to get them to fill out the weekly log if they had put it aside, so the response rate for the log represents a "single wave" response rate. (For the productivity portion of this study, we were able to follow up in a more traditional fashion. However, our overall response rate was not much different for that part of the study than for the weekly workload questionnaire.) Table 1 shows the response rates for each week of the study.

Table 1. Response rates by week for weekly diary

Week	Number Responding	Response Rate (%)
1	16	67
2	13	54
3	19	79
4	17	71
5	20	83
6	20	83
7	17	71
8	19	79
9	21	88
10	15	63
11	16	67
12	17	71
13	16	67
14	14	58
15	13	54
16	14	58
Overall	267	70

We were pleasantly surprised with the response rate that is reflected in the above study,

considering the constraints placed on the kind of follow-up that could be implemented.

Estimation Procedures

The method of getting estimates would have been very straight-forward in the absence of non-response. The proper weights for the stratified design were easily obtained and the computational scheme would follow easily from that. However, since there was only one respondent per stratum in a given week, a collapsed-stratum approach or some other method of variance estimation would have had to been implemented in order to provide a measure of the error in the estimates.

With the non-response problem taken into the picture, some weighting adjustment for the non-responses would have to be implemented. (There was no problem with non-coverage or with some of the other problems that typically call for a weighting adjustment in a sample survey.)

The general weighting adjustment goal was to obtain a set of weights such that the sum of the weights equals the size of the population. (In our case, the sum of the weights across all sixteen weeks would equal 16 times the population size.) Because of the nature of the design, a usual weighting adjustment for non-response created difficult problems in coming up with a set of compatible weights. Therefore, we opted for using an imputation method for the non-responses and then applied the regular weights to the set of responses so obtained. Our method of imputation was essentially equivalent to a hot-deck method. For those strata that were empty because of a non-response, a respondent who was identical or as similar as possible on the basis of four demographic variables was imputed into that particular stratum. This imputation process was performed week by week, using only those respondents from that week's set of responses for the missing cells. After this imputation was carried out through the entire 16 weeks' samples, then ordinary estimation formulas for a stratified sample were applied.

To get estimates of the standard errors we took advantage of the 16 independent samples we had in place of using a collapsed-stratum approach or other variance estimation procedure. Assuming that there was no difference in weeks for the various measures of interest, the variance of the 16 week's estimates could be used to provide a measure of the error of estimate using a "replicated sampling" approach. This provided a much simpler variance estimation procedure than any other. (If the assumption of no-difference in weeks is false, then the standard errors obtained by this approach would be an over-estimate of the standard error and would simply place us a conservative position in reporting standard errors or computing confidence intervals.) Table 2 presents a few of our estimates and their estimated standard errors.

Table 2. Some estimates of workload and their standard errors

Description	Estimate (hr's.) (Average Week)	Standard Error (hr's.)
Total Hours/week	51.55	0.88
Monday	9.36	0.35
Tuesday	10.15	0.14
Wednesday	9.88	0.16
Thursday	9.35	0.35
Friday	8.59	0.22
Saturday	4.23	0.23

Conclusions

In retrospect, we feel that we have succeeded in our initial goal of getting more precise estimates than traditionally would be obtained for studies of this type. The degree of faculty cooperation in keeping a weekly log of time spent, day-by-day, would encourage us to use the same approach if asked to do this again. The simplicity of the design and the estimation strategy was straight-forward, with no cumbersome formulas needed to get estimates and their standard errors.

Of course, there are still problems that need to be corrected, and some problems which may never be overcome. We wonder if we might have used a rotating design similar to that used in the CPS which would allow us to measure week to week changes as well as give us a good estimate of the typical work week of faculty. Answers to such questions will have to wait for another study.

References

Cochran, William G., (1977), *Sampling Techniques*, (3rd ed.), New York, John Wiley & Sons.

Levy, P. S., Lemeshow, S., (1991), *Sampling of Populations: Methods and Applications*, New York: John Wiley & Sons.

McCarrey, Leon R. (1978), *Faculty Activity Survey-State Universities, Colleges, Community, and Technical Colleges*, Office of Academic Affairs.

Russell, S. H., Cox, R. S., Williamson, C., Boismier, J., Javitz, H., Fairweather, J., and Zimbler, L. J. (1990), *Faculty in Higher Education Institutions, 1988*, Washington, D. C.: Survey Report of National Center for Education Statistics.

Utah System of Higher Education. (1991), *USHE Faculty Workload Report*, Salt Lake City, Utah: State Board of Regents.