

- 4) the answer is wholly incorrect or no answer is given, and
- 5) no attempt is made to answer the question.

This system allowed Childers flexibility in scoring. He could collapse categories in any combination and then assign numerical values in order to test difference scales of scoring (Childers 1971: 116-117). In order to maintain the objectivity of scoring, his study would only look at “requests for simple factual answers” (Childers 1971: 81, 103).

In 1969, two years after Crowley’s investigation, Childers performed the first analysis of reference service using multiple regression. This statistical method enabled him to analyze the cumulative effects of more than one independent variable on reference accuracy. For the first time, the theory espoused in Goldhor’s research was tested using numerous independent variables to measure both staff ability and library collection. Unfortunately, Childers also examined some variables, such as hours of operation and circulation per capita, for which no theoretical relationship to reference service was advanced. Both of these variables were included in the final regression formula which explained 89.83% of the variance in the dependent variable (Childers 1971: 157-159). However, regardless of the equation’s high predictive power, the lack of theory prevented the ability to draw any causal inferences. Notwithstanding these criticisms, Childers did advance the level of statistical sophistication being used for analysis of reference service. Multiple regression would be used in four subsequent studies (Jirjees 1983; Myers 1983; Benham 1987; Powell 1987).

Of these four studies, perhaps the most important was the first conducted in 1975 by Ronald Powell who attempted to test the effects of reference collection size on reference performance. His review of the literature concluded that little statistical evidence existed to support the conventional theory that collection size had a significant effect on reference accuracy

(Powell 1987: 166, 170). Powell, like Bunge, used the librarian rather than the library as the unit of study for his research. Consequently, Powell was able to test the cumulative effects of both librarian characteristics (e.g., education and experience) and library characteristics (e.g. collection size and budget) in his analysis. He determined that reference collection size had a curvilinear relationship to accuracy in that the effect size diminished as collection size exceeded 3,500 volumes (Powell 1987: 238, 251-252, 257). He further concluded that variables measuring librarian characteristics had a strong association with performance (Powell 1987: 226-234, 258-259). The findings substantiated the theory that accuracy was dependent on the library collection and the ability of staff.

Since the time of Childers's project in 1969, twenty-five studies of reference accuracy have been conducted that did not use multiple regression, including two in which Childers himself was an investigator. In one instance, a Pearson's product moment correlation was used (Lowenthal 1990). In three instances, a chi-square test of significance was administered (Ramsden 1978; Hernon and McClure 1982; Way 1987). In all other instances, findings were reported as simple percents of the number of questions answered correctly on a reference test (King and Berry 1973; House 1974; Peat, Marwick, Mitchell and Associates 1975; Childers 1980; Schmidt 1980; DeWath 1981; Weech and Goldhor 1982; Wise 1982; Yellot and Barrier 1983; Rubinstein 1984; Van House and Childers 1984; Gers and Seward 1985; Birbeck and Whittaker 1986; Hansel 1986; Hernon and McClure 1986a; Hernon and McClure 1986b; Rodger and Goodwin 1987; Williams 1987; Lea and Jackson 1988; Christensen et al 1989; Head and Marcella 1993). While many descriptive characteristics of libraries were discussed, none were tested for association with reference performance. However, this body of research did serve to support, or perhaps "entrench," the fifty-five percent rule.

## B. Tests of Satisfaction

As an alternative to accuracy, another measure of reference service which has been explored is user satisfaction. This vein of research tends to examine the effect of variables over which librarians have more direct control, such as behavior and communication abilities, rather than variables which are largely imposed on the librarian as part of the reference environment, such as collection size. Consequently, the results of these studies might have more immediate benefit to practitioners in the field. However, because of the greater complexity of the measures which are involved, significant findings have been harder to obtain. Notwithstanding this difficulty, research into user satisfaction has certainly contributed to heightening awareness within the profession regarding the theoretical variables that contribute to increased reference performance.

The first investigation into user satisfaction with reference service was performed by the Enoch Pratt Public Library through a telephone survey of a random sample of readers who had asked questions at the reference desk. The findings indicated that readers were generally satisfied with the service, with only two out of thirty-seven readers indicating that service quality was poor. The accuracy of information being provided to the reader was never examined. The purpose of this study was not to determine any statistically valid measure of satisfaction, nor to test an instrument for gauging satisfaction, but merely to acquire some evidence regarding whether or not users were pleased with the service (Enoch Pratt Free Library 1968: 1-3).

In 1976, Helen Gothberg conducted an experiment to determine the effect of verbal and nonverbal communication on user satisfaction with the reference process. Specifically, she examined the role of immediacy, the quality of liking or closeness in an interpersonal relationship,

on three different dimensions of satisfaction: the user's satisfaction with reference interview, the user's satisfaction with their own performance in the interview, and the user's satisfaction with the transfer of information between the librarian and the user (Gothberg 1976: 127). Using analysis of variance on data gathered from the observation of real reference transactions and a survey of real inquirers, she concluded that librarians who exhibit immediate communication skills will increase the amount of user satisfaction (Gothberg 1976: 129). While the study was firmly based on theory and the method was appropriate for comparing two conditions, data was gathered on only two librarians who conducted fifteen reference transactions each. The small sample decreased the confidence of the findings, and cast doubt when subsequent studies with larger samples found no significant differences.

In 1983, Marilyn J. Markham, Keith H. Stirling, and Nathan M. Smith conducted a study based on Gothberg's method, although the focus of their effort was determining the effect of self-disclosure, rather than immediacy, on user satisfaction. Self-disclosure is the extent to which an individual shares feelings of beliefs with another, and thereby increasing the likelihood that the other person will respond candidly. Although they increased the number of reference staff from two to four and the number of interviews from fifteen to sixteen each, the findings of their study was subject to the same limitations which affected Gothberg's investigation. Markham, Stirling, and Smith concluded that self-disclosure had no significant effect on the outcome measure, although some significant differences did appear when items on the instrument were examined individually (Markham, Stirling, and Smith 1983: 371). This finding suggested that the instrument itself required further testing and revision before conclusive findings could be obtained.

That same year, George D'Elia and Sandra Walsh conducted an analysis of user satisfaction with different forms of library service. Although reference service is not explicitly

examined, variables concerning the openness and helpfulness of library staff were included. Their study utilized a sample of 523 library users which was the largest sample taken for service evaluation at that time. The goal of this investigation was to test a construct for measuring user satisfaction which had not been rigorously tested (D'Elia and Walsh 1983: 115). Multiple regression was used to measure the effects of numerous variables on the construct. In reporting the findings, this study demonstrated a high methodological standard by discussing the content and construct validity of the measures and providing full descriptive statistics for each test. In conclusion, D'Elia and Walsh determined that the construct was not effective for describing relationships between user satisfaction and public library services.

“User behavior both in terms of evaluation and use of the library appears to be enigmatic. In the final analysis, this study demonstrates the complex nature of user behavior and our limited understanding of such behavior.” (D'Elia and Walsh 1983: 132).

This statement articulates the complexity of the reference transaction that unconsciously argues the necessity for complex, multivariate methods of assessment.

In 1986, Roma M. Harris and Gillian B. Michell attempted a more controlled investigation of user satisfaction by moving the study from the reference desk to a laboratory setting. A large sample of 320 public library users were asked to watch a video of a staged reference transaction and to answer questions about the behavior of the librarians. Using analysis of variance, the responses indicated that social variables such as gender, warmth of the librarian, and inclusion (i.e. the extent to which a librarian adopts a teaching role) have an effect on reference outcomes (Harris and Michell 1986: 94-99). While this investigation did not serve to evaluate performance for any given institution, it did propose a method for increasing the understanding of the reference process and for devising valid measures which could be used in later evaluation.

In 1988, Marynelle DeVore-Chew, Brian Roberts, and Nathan M. Smith performed a study to examine the effects of nonverbal communication on user satisfaction. While the method was similar to the earlier effort by Gothberg, the sample of 354 library users in this study was much larger. For the first time, multivariate analysis was used to examine reference service outcomes, although it was not effectively applied. No significant findings were observed. The investigators suggested that the sample was not large enough to account for differences in the population because of the high number of variables which were being examined, and thus the findings may be an example of Type II error in which an incorrect null hypothesis is accepted (DeVore-Chew, Roberts, and Smith 1988: 397-398).

The largest samples which have ever been drawn for the purposes of reference service evaluation all employed user satisfaction as the outcome measure (see Table 1) (Bunge 1990: 42; Childers 1996; D'Elia and Rodger 1996: 292).

Table 1: Largest Samples in Reference Evaluation Research

<b>Study</b>	<b>Sample Size</b>
Wisconsin-Ohio Reference Evaluation Program	4,800
California Reference Evaluation Program (CREP)	13,000
D'Elia and Rodger	25,684

In these studies, after a reference transaction is completed, genuine inquirers are asked to fill out questionnaires regarding their satisfaction with the reference process. Whereas the Ohio-Wisconsin and California studies also attempt to measure accuracy, both studies employ the user's perception of accuracy (i.e., did you find the information you wanted?) as the outcome measure. This subjective measure is not comparable to the objective measures used in traditional accuracy studies.

The large samples in these studies produce high levels of statistical power, reducing the risk that the findings suffer from either Type I or Type II error. However, these studies suffer from threats to content validity in that the sample does not cover the total domain of library users. All three studies rely on users who are willing to fill out survey questionnaires. This practice results in self-selection bias since it is probable that those users who have good experiences are more willing to take the time to fill out the survey. Of additional concern is the fact that library staff are probably reticent to offer a survey to those users who are unpleasant or express dissatisfaction with service. Consequently, the subjects are not randomly selected. No attempt was made in any of these three studies to determine to what extent the sample was representative of the total population.

In 1990, Jo Bell Whitlach conducted the most sophisticated study of reference service which has been performed up to the present time. Proceeding from a strong theoretical base, Whitlach tested a model of reference service that identified three service outcomes: the librarian perception of service value, the user perception of service value, and the success in locating needed materials (Whitlach 1990b: 23). Three stepwise multiple regressions were performed, one for each dependent variable, to identify strong predictor variables for performance. The findings indicated that a strong service orientation on the part of librarians and direct “feedback” from readers during the reference transaction contributed to higher levels of reference performance (Whitlach 1990b: 26-29).

The measures for independent variables Whitlach used to predict service outcomes were originally developed for her study, but many were based on established measures from other disciplines (e.g., the Internal Work Motivation scale from the Michigan Organizational Assessment Questionnaire) (Whitlach 1990b: 72). Her research benefited greatly from the use of

7-point Likert scales which were capable of registering a wide range of variation in the sample for any given variable. Measures for tested for internal reliability prior to performing regression analysis. This strong concern with methodological issues exhibited by Whitlach serves as an example to others and sets a high standard for subsequent researchers to follow, including the present investigator. The current study emulates Whitlach in many characteristics, but also grows beyond her efforts through the application of HLM.

### 3. MULTIVARIATE ANALYSIS

Murfin and Gugelchuk were the first to utilize cluster analysis in comparing thirty-five variables related to the reference transaction (Murfin and Gugelchuk 1987: 324-325). While this method can not be considered a true multivariate method because it does require that dependent variables be identified, cluster analysis is a precursor to other forms of multivariate analysis, such as factor analysis or canonical correlation. Murfin and Gugelchuk thus raised the level of statistical complexity being used in reference service evaluation beyond the sole reliance on univariate procedures such as multiple regression and analysis of variance.

As mentioned earlier, the first multivariate study in the area of reference service evaluation was performed by DeVore-Chew, Roberts, and Smith. They used the method of multivariate analysis of variance, although their sample was not large enough to use the method effectively (DeVore-Chew, Roberts, and Smith 1988).

All other studies to the present have employed univariate statistical procedures to measure reference service. In cases where more than one dependent variable was examined, each variable was tested separately using a univariate method. Since the time of these initial investigations,



advances in computer software have made it more feasible for investigators to utilize multivariate methods.

#### 4. OBTRUSIVE VERSUS UNOBTRUSIVE OBSERVATION

Several researchers have suggested that unobtrusive observation of reference transactions will produce data that more closely resembles true field conditions than obtrusive observation (Childers 1971: 120; Crowley 1971: 25; Myers 1979: 15). Librarians who are aware they are being tested have greater motivation to perform well. Because of the logic and prevalence of this belief, over half of the reference evaluation studies which test accuracy have been unobtrusive. Unfortunately, the practice of unobtrusive observation limits the investigator's capability to probe into the opinions and attitudes of librarians and users. Consequently, all of the studies which examine user satisfaction employ obtrusive methods.

To what extent does obtrusive observation bias findings? Weech and Goldhor determined that reference departments evaluated using obtrusive methods answered fifteen percent more questions correctly than those evaluated using unobtrusive methods, yet also determined that the statistical relationship between the method of evaluation and the results of evaluation was slight (Weech and Goldhor 1982: 316-319). However, this finding is based on a sample of only five libraries. A larger sample may have resulted in a greater value for that relationship.

This evidence seems to indicate that obtrusive observation will bias findings towards describing higher levels of performance than might actually occur under field conditions. Nonetheless, the use of obtrusive observation is essential in order to gain personal information from human subjects. This personal data can be used to help investigators develop models of the reference process. Once this process is more clearly understood, sophisticated measures of service

outcomes can then be devised which might enable unobtrusive observation of satisfaction and utility. Such measures could include recording the visible, non-verbal indicators of satisfaction and interest in the reference transaction, or surveying library users without the knowledge of the librarians.

## 5. SAMPLING PRACTICES

Random sampling is the strongest defense against bias in any study design. While many studies in the area of reference service evaluation claimed to utilize random sampling, it has rarely been executed. As mentioned earlier, the largest samples ever drawn for evaluation research suffer from self-selection bias. In addition, numerous studies that sample from more than one library fail to account for hierarchical effects. Sometimes, libraries have been selected randomly, yet individual librarians within them have not.

Many studies take place within a single library system, or even within a single library. In such studies, the entire target population is often included in the sample. This type of sampling may be appropriate for the goals and boundaries of the given investigation, but the findings have little applicability to other institutions.

Of additional concern to the lack of random sampling is the prevalence of low sample sizes. Of 59 studies performed between 1967 and 1993, only 21 had more than 50 subjects, and 14 had 12 or fewer subjects (Saxton 1997: 283-288). Determining the preferred sample size for any study is affected by two concerns. First, as the number of variables being studied increases, the size of the sample should be increased. Second, as the magnitude of variance for a given parameter within a population increases, the size of the sample should be increased. In the history of reference service evaluation, neither concern seems to have received a large amount of

attention. Given the complexity of the subject, the number of variables being examined in most studies, and the lack of knowledge concerning the variance of parameters, one would expect to see average sample sizes in the hundreds as a minimum.

## 6. CAUSES OF REFERENCE FAILURE

While research presented in the literature clearly suffers from many methodological shortcomings, much anecdotal evidence has been acquired regarding what contributes to quality reference service. When the question is examined conversely, several reasons are suggested as causes for reference failure.

First, in numerous studies the prime reason suggested for not being able to answer a question was the librarian's failure to probe adequately (Halldorsson and Murfin 1977: 394; Childers 1980: 926; Wise 1982: 27; Gers and Seward 1985: 33; Rodger and Goodwin 1987: 141-142; Lea and Jackson 1988: 582; Christensen 1989: 470). Too often, librarians will take a question at face value without attempting to determine the actual information need. Wise noted that when librarians were asked for information about a given individual, two-thirds of those queried automatically assumed that the library user wanted short biographical information and began searching for an answer without any further question negotiation. Some studies have been specifically designed to study this phenomenon, using what are loosely termed "escalator questions" (Halldorsson and Murfin 1977, 387; Childers 1980, 925-926). These are questions which are initially presented to the librarian with missing or faulty information worded in such a way as to misrepresent the actual test question, requiring the librarian to probe in order to achieve success.

Second, the failure to make referrals is another common cause attributed to reference failure. In some studies, librarians exhibited an apparent reluctance to make referrals to other agencies (Wise 1982: 27; Myers 1983: 109; Hernon and McClure 1986: 40-41; Christensen 1990: 472). This failure reduces accuracy rates because some studies do score higher points for making a referral than for giving no answer. Such a finding is surprising when one considers that

between Gale's *Encyclopedia of Associations* and the *United States Government Manual*, the librarian has over twenty thousand agencies and organizations from which to choose.

Third, the failure to conclude the reference transaction with any kind of follow-up designed to determine if the library user has received the answer they were after has a substantial impact on reference accuracy (Gers and Seward 1985: 34). Librarians may be hesitant to risk uncovering more complex questions from a seemingly satisfied library user, particularly when the desk is busy or it is close to closing time. However, all librarians should conclude each reference transaction with the simple question, "Did you find what you wanted?" If such a practice were universally implemented, it is still not a guarantee that the librarian would then be able to produce an accurate answer to the question, but would prevent a library user from leaving the library unsatisfied while the librarian is under the impression that the information need has been met.

## CHAPTER III METHODOLOGY

This chapter will begin with a discussion of the assumptions that must be met in conducting hierarchical linear modeling and how this study will address the common threats to validity. Next, the target population will be identified, followed by an explanation of the sampling methods that will be used. Afterwards, a technical description of each variable will meticulously explain how each parameter is to be measured. The chapter concludes with an organizational plan outlining the steps involved in conducting the experiment.

### 1. ASSUMPTIONS AND THREATS TO VALIDITY

Hierarchical linear modeling using ordinal variables requires that the following assumptions be met (Bryk and Raudenbush 1992: 200):

- 1) All relevant variables have been included in the model.
- 2) All independent variables are fixed and measured without error.
- 3) The dependent and independent variables have a linear relationship.
- 4) The independent variables are uncorrelated.
- 5) The error (i.e.,  $e_{kij}$ ) for each observation is independent and normally distributed with a mean of 0 and constant variance for every transaction (i.e., the first level unit) performed by each librarian (i.e., the second level unit) within each library (i.e., the third level unit).
- 6) The error (i.e.,  $e_{kij}$ ) is not correlated with the independent variables.
- 7) The vector of random errors at the second and third level exhibit multivariate normality.
- 8) The set of second level independent variables is independent of the second level error (i.e.,  $u_{ki}$ ).

- 9) The set of third level independent variables is independent of the third level error (i.e.,  $U_k$ ).
- 10) The errors at first, second, and third levels are independent.

In addition to addressing the assumptions listed above, all data will be screened before the regression analysis is performed. Each variable will be examined for normality to determine whether assumptions have been violated. Outliers, which will be defined as any value greater than 2 standardized residuals above the mean (Pedhazur 1982: 37-38), will be deleted from the sample. Cases with missing data will be deleted from the sample. However, should this process result in having to remove a large number of cases (e.g., ten-fifteen percent of all cases) then the conventional method of replacing missing data with the mean for that particular parameter will be utilized.

The literature review identified many threats to validity in previous research. To neutralize these threats, six elements have been included in the design of this study:

- 1) obtaining data on a large sample ( $N \geq 10,000$ ) to reduce the variance of each dimension and determine findings that are highly generalizable,
- 2) using random sampling of sites to minimize bias,
- 3) assessing the degree to which the sample is representative of the total population,
- 4) minimizing the distrust of participating librarians to reduce bias by guaranteeing confidentiality and inviting them to help design portions of the instrument unique to their institution,
- 5) achieving high participation among librarians to minimize selection bias by engaging enthusiasm for successful completion of the project and promising valuable results, and
- 6) achieving high participation among library users to minimize selection bias by designing a survey instrument that can be completed easily and quickly.

These efforts should minimize the effects of bias and help strengthen arguments against potential criticism regarding the validity of the study findings.

## 2. THE POPULATION, SAMPLING, AND SAMPLE SIZE

The target population of this study is all reference transactions that occur in public libraries. According to the National Center for Education Statistics, this population totaled approximately 242.7 million transactions in fiscal year 1993 (National Center for Education Statistics 1995: 30).

All 76 public library jurisdictions in the Los Angeles basin will be invited to participate in the study. Of those who express interest, 20 will be randomly selected using a stratified sampling method based on library size in order to obtain a representative sample of sites. The sampling of reference transactions will occur during a three-week period. At each library, sampling will occur over a period of thirty hours. Data collection will occur for one hour at a time. Each hour-long timeslot will be randomly staggered throughout the three-week test period. All reference queries asked during every hour-long timeslot will be recorded. According to data from the California State Library, this sample will consist of approximately 10,000 reference transactions (California State Library 1997). Assuming that at least half of the users who initiate these queries respond to the survey, which is a conservative estimate when compared to the CREP response rates, then data from a minimum of 5,000 transactions will be available for multiple regression analysis, which is more than adequate to achieve the statistical power necessary to determine significant findings (Bryk and Raudenbush 1992: 211).

Reference transactions will be recorded by writing down the negotiated query, the answer that was given, and the source. Although the task seems labor intensive, similar techniques have

been used successfully in the past (Guerrier 1936; Van Hoesen 1948). Writing down the question and answer arguably takes less time than answering a fifteen item questionnaire about the reference transaction. If the volume of traffic at the reference desk grows so busy that it becomes unfeasible to record every query, librarians will be asked to record as many as they can and to count the remainder. In this fashion, the investigator will be able to determine what proportion of queries during the sampling period were captured and how many were unrecorded.

In addition to enabling reference service evaluation in this study, subsequent analysis of the data will be used to calculate many descriptive statistics which are outside the boundaries of this investigation but will be of value to the participating libraries, such as determining how frequently certain reference sources are being used and determining which subjects are of greatest interests to inquirers, and determining collection strengths and weaknesses. These statistics can be used to provide evidence for decision-making in collection development, training new staff, and preparing grant applications. A report containing these descriptive statistics is one of the tangible products which is offered to libraries as an incentive to participation. This data will also create a large archive of genuine reference queries that could be used for subsequent study design or the testing of new reference tools.

The librarian receiving the query will determine whether or not it is a ready-reference or a research query. Although the boundary between the two query types is not clearly demarcated, Hutchins proposed a general distinction. She wrote that a ready-reference query requires locating information in a published source that has already addressed the question, whereas a research query requires the reader to collect and analyze information in order to draw his or her own interpretations. She specifically criticized using the measure of time it takes to answer the question as a means of distinguishing the two (Hutchins 1944: 16-17). Bereft of any mechanical



means for labeling queries, this study will rely on the judgment of participants. These random individual judgments spread over a large number of transactions will prevent any systematic error from skewing the data.

To facilitate the recording effort, some ready-reference questions will be identified as frequently asked questions (FAQs). A FAQ is a question that is asked so frequently that the query resolution has become routine. Each library has its own unique set of FAQs which arise from the most common needs of the service population and the special strengths of the library collection. Reference librarians at each site will be surveyed to determine the top five most frequently asked questions for their library. Once these queries have been identified, they will be enumerated on the transaction instrument and librarians will record them by simply “checking off” the category. FAQs will be tabulated to determine what proportion they represent of the total sample.

Other encounters that are not true reference queries will also be counted to determine their impact on reference staff. Directional queries will be simply counted. Questions posed by the homeless or persons under the influence of a controlled substance may at first seem like a reference query, but might be reduced to nonsense as the transaction progresses. These encounters will also be simply counted. By categorizing not only reference queries but also their entire interaction with the public in this manner, reference staff will stay focused on the goal of recording each and every transaction. It will also prevent librarians from initially dismissing queries posed by the homeless or other disenfranchised persons when they later prove to be genuine reference transactions.

### 3. VARIABLES

As discussed in the first chapter, all the variables in this study will be measured on an ordinal 7-point Likert scale. The seven degrees on the scale limit the influence of range restriction by allowing a wide variety of response. Using a 7-point scale for each variable will also facilitate the interpretation of standardized scores.

#### A. Dependent Variables

This study will assess reference service performance in terms of three variables: accuracy, satisfaction, and utility. Past research supports the idea that accuracy and satisfaction are unrelated, but this situation is probably the result of range restriction in the case of measuring accuracy as a bivariate variable (i.e. accurate = 1, inaccurate = 0). Logically, better service should result in better communication that would increase the degree of accuracy of the information being given to the user, as well as increasing the probable utility of that information. Allowing additional degrees of response in the measurement of accuracy may reflect a relationship that has been heretofore hidden.

These three variables are scaled as follows:

#### **Y<sub>1</sub> - Accuracy**

- 7 Wholly accurate - A complete answer, and all information is correct**
- 6 Partially accurate - An incomplete answer, but the information provided is correct**
- 5 Accurate Referral - User is referred to an agency that can supply the correct answer**
- 4 No Answer**
- 3 Partially inaccurate - User is given both accurate and inaccurate information**
- 2 Inaccurate referral - User is referred to an agency that cannot supply the correct answer**
- 1 Inaccurate answer - No correct information is given**

The rank order indicates increasing degrees of accuracy and information being provided. On this scale, a complete answer is scored higher than a referral because more information is being provided to the user. A “No answer” is scored higher than an inaccurate answer because less inaccurate information is being provided to the user. An inaccurate referral is scored higher than an inaccurate answer because that user might ultimately be directed to an appropriate agency and obtain the accurate answer, whereas providing a wrong answer does not leave the user with any other option to pursue.

This scale is a modification of Childers’ design for assessing the accuracy of telephone queries (Childers 1971: 115-117). According to the Childers’ scale, distinction is made between a librarian refusing to answer a question and a librarian searching for an answer but finding none and making no reply. In this scale, both outcomes are categorized together as a non-answer because the response to the user is the same. Also, because it was only intended for ready-reference queries, Childers’ scale did not account for the role played by referrals in answering queries.

## Y<sub>2</sub> - User Satisfaction

Users will indicate one of the following values in response to the question, “Are you satisfied with the service you received from the librarian?”

1	2	3	4	5	6	7
Not Satisfied			Somewhat Satisfied			Very Satisfied
Poor Service			Adequate Service			Excellent Service

This question is a variation of the type used by Whitlach. In order to encourage more users to fill out the survey, the question asked here has been rephrased to appear less clinical than, “Indicate how satisfied you are on the following scale” (Whitlach 1990b: 57 and 67).

### Y<sub>3</sub> - Utility

Users will indicate one of the following values in response to the question, “Are the materials you found useful for your research?”

1	2	3	4	5	6	7
Useless			Somewhat Useful			Very Useful

This construct will only measure the immediate utility that the user perceives from brief inspection of the materials they have been provided by the librarian. This trait might be more descriptively named “short-term utility.” Ideally, measuring utility would occur at some later point after the user had enough time to work with the information they received. However, measuring “long-term utility” would require gaining the consent of users to be contacted and obtaining address and/or telephone numbers. At present, short-term utility may be the only quality that can be feasibly measured, and would logically have a high correlation with long-term utility. In the future, a study whose object is to correlate short-term and long-term utility would indicate how well the first measure could predict the second.

Whitlach measured some facets of utility when she asked users whether they had received enough information and also asked them to evaluate the degree of relevance the materials had to the query and (Whitlach 1990b: 58 and 67). In this study, the amount of information has been included as an aspect of accuracy in terms of whether the user has received a complete or partial

answer. Relevance is an imprecise term that may confuse many public library users (and librarians). In this study, the question of usefulness is asked directly and simply.

## B. FIRST LEVEL INDEPENDENT VARIABLES

The study will examine nine different variables as first level predictors of reference performance. These variables will be measured as follows:

### B.1. The Difficulty of the Question

What makes one question more difficult than another? Is the degree of difficulty merely a situational condition dependent on the librarian's own knowledge state? Or does a question contain identifiable elements that make it more difficult to answer? In reviewing the literature, Childers, Lopata, and Stafford concluded that no standard definitions of query difficulty exist (Childers, Lopata, and Stafford 1991: 238-239). To investigate these questions, difficulty will be measured in this study from two perspectives: complexity of the question and the currency of the query.

Who is the best judge of difficulty? Often, a reference transaction is a difficult process not because of the query but because of the personality of the inquirer or the library user's lack of research skills. A librarian's assessment of the query is likely to be biased by the difficulty experienced in working with the user. To separate user traits from query traits, difficulty will be assessed by outside observers.

The judges panel will rate complexity. This variable will be measured as follows:

### X<sub>1</sub> - Complexity

How complex is the query?

1	2	3	4	5	6	7
Not complex			Somewhat complex			Very Complex
Query involves one topic					Query involves numerous topics	

This measure is originally devised for this study.

Current awareness has long been recognized as a significant problem leading to reference failure (Crowley 1971: 54-56). In this study, the investigator and two research assistants will assess currency since this quality can be mechanically operationalized

### X<sub>2</sub> - Currency of the Query

7	Within this week
6	Within two weeks
5	Within this month
4	Within three months
3	Within six months
2	Within one year ago
1	More than one year ago

This scale represents currency as a property that decreases exponentially as time passes. In other words, a query regarding something that occurred six years has about the same degree of currency as something that occurred eight years ago. However, something that occurred this past week has a far greater degree of currency than something that occurred last month.

In addition to determining difficulty, queries will also be examined to determine their nature, or type. The classification scheme developed by Van Hoesen will be used to sort queries into different categories (Van Hoesen 1948: 64-67). This characteristic will not be used in the structural equation model since the data is categorical rather than ordinal. Furthermore, complexity and currency are unrelated to the nature of the question. For example, a librarian could receive statistical queries that are either simple or complex. Biographical questions could be

historic or pertain to someone who appeared in the news the previous evening. Nonetheless, this descriptive data will be of value to the participating libraries and obtaining it serves as additional incentive for participation.

## B.2. The Experience of the User

In the reference process, the library user assists in discovering the resolution to the query (Whitlach 1990a: 208-209; Brown 1994: 213-214). The inquirer works with the librarian to determine the exact query. The inquirer provides feedback to the librarian regarding whether they can comprehend a particular source. Clearly, the user is in part responsible for a successful outcome. Logically, user characteristics that indicate the ability to interact with the librarian and to use a greater proportion of the library collection are likely to influence service performance.

Users who visit the library frequently are more likely to feel at ease and more open to revealing questions they might have to the staff who works there. They are more likely to view the institution as a place to seek help, and are also more likely to have had good experiences at the library. Likewise, people who use reference service frequently are more likely to feel at ease with librarians, are more likely to be familiar with the reference process, and are more likely to have a greater degree of trust in working with librarians based on good experiences in the past. This familiarity contributes to better service performance because the experience of the library user reduces potential anxiety and encourages openness. Familiarity will be measured in two ways as follows:

X<sub>3</sub> - Familiarity with the Library

How often do you use the library?

1	2	3	4	5	6	7
First Time	Now and Then	Monthly	Once every Two Weeks	Weekly	Twice a Week	Three Times a Week

X<sub>4</sub> - Familiarity with Reference Service

How often do you ask questions at the reference desk?

1	2	3	4	5	6	7
First Time	Now and Then	Monthly	Once every Two Weeks	Weekly	Twice a Week	Three Times a Week

The first is taken from the study performed by Whitlach (Whitlach 1990b: 68). The second measure specifically measures familiarity with reference service and the response values are identical to the first measure.

Logically, persons with greater educational attainment are more likely to be able to comprehend a wider variety of library materials, both in terms of breadth of format and depth of knowledge. Educational attainment will be measured here as:

X<sub>5</sub> - Level of Education

What is this highest level of education have you have achieved?

1	2	3	4	5	6	7
Some High School	High School Graduate	Some College	Associate Degree	Bachelors Degree	Some Graduate School	Graduate Degree

This measure is derived from the *United States Census*. However, while the census specifically delineates between the attainment of a masters degree or a doctorate, this study will not because persons with such a high level of education are a very small proportion of the service population for public libraries. Assigning a value for each graduate level would needlessly skew the distribution of scores and constrain the amount of variation within the population that could be assessed by the measure.



### B.3. The Quality of Communication

In 1996, the Reference and User Services Association (RUSA) published a set of standards outlining behavior which is generally accepted as contributing to improved performance (RASD Ad Hoc Committee on Behavioral Guidelines for Reference and Information Services 1996). These behaviors are concerned with optimizing the quality of communication between the librarian and library user. These standards have been operationalized here in four questions that will be addressed to library users. This study is the first attempt to test these standards empirically.

These variables are measured as follows:

X<sub>6</sub>

Did the librarian welcome you and appear ready to assist you?

1                    2                    3                    4                    5

Librarian appeared  
not to see me

6                    7

Librarian smiled and  
offered to help me

X<sub>7</sub>

Did the librarian appear interested in your question?

1                    2                    3                    4                    5

Librarian appeared  
distracted or bored

6                    7

Librarian was focused  
on my question

X<sub>8</sub>

Did the librarian attempt to understand your question?

1                    2                    3                    4                    5

Librarian interrupted me  
and didn't listen

6                    7

Librarian repeated the  
question to verify it

X<sub>9</sub>

Did the librarian attempt to verify you had located what you wanted?

1                    2                    3                    4                    5

Handed me a book  
and moved on

6                    7

Asked me if I had  
found what I wanted

### C. SECOND LEVEL INDEPENDENT VARIABLES

The characteristics of individual librarians comprise the second level in the hierarchical design of this study. The ability of the librarian has frequently been theorized to affect reference performance. Measures of education and job experience have often been used to operationalize ability. In recent studies, job satisfaction has also been used. This study will examine the effects of all three.

#### X<sub>10</sub> - Reference Service Experience

The librarian's experience is a characteristic that has been examined in seven previous studies, yet quantification has always proved problematic. Frequently the number of years an individual has worked in libraries does not accurately reflect the amount of time that person has spent in reference work, nor does it distinguish between full-time and part-time employees. Arguably, many other professional tasks, such as collection development and cataloguing, all contribute toward better performance at the reference desk, yet probably not as effectively as actual reference service experience. Likewise, experience in other service professions will also contribute to one's performance. If a researcher attempts to quantify total library work experience, or total work experience regardless of setting, extra weight should be given to reference experience over other types.

For this study, experience will be limited to reference desk experience measured as an estimate of total hours the librarian has spent on desk. The numerical values will be categorized in seven steps as follows:

7	24,001 or more hours	26+ years
6	20,001 - 24,000 hours	22 years
5	16,001 - 20,000 hours	18 years
4	12,001 - 16,000 hours	14 years
3	8,001 - 12,000 hours	10 years
2	4,001 - 8,000 hours	6 years
1	1 - 4,000 hours	2 years

The approximate years of experience are based on a librarian who worked an average of 20 hours on the reference desk per week each year.

#### X<sub>11</sub> - Education

Historically, research on the effects of a librarian's education has followed two lines: comparisons between librarian with and without and MLS and comparison between librarians with and without expertise in a second discipline as evidenced by multiple graduate degrees. The former comparison is less viable today since a majority of librarians do possess masters degree in library science. The second comparison is perhaps of increasing importance as information becomes more specialized. The following measure accounts for these concerns:

#### Education

7	Graduate degree in LIS and second graduate degree
6	Graduate degree in LIS and some other graduate work
5	Graduate degree in LIS
4	Other graduate degree, some LIS education
3	Other graduate degree
2	Some graduate work
1	No graduate work

This measure has been originally devised for this study. This variable does not take into account the effect of continuing education programs beyond a LIS graduate degree that many librarians pursue, whether they be workshops, night courses, professional association events, or simply professional reading.



Factor analysis will be used to determine which variable is the strongest indicator of satisfaction.

#### D. Third Level Independent Variables

The characteristics of individual libraries comprise the third level in the hierarchical design of this study. In reference service evaluation, collection size has been the most frequently used measure of the library environment. In addition to collection size, this study will examine the effects of service policy and managerial support given to reference staff.

#### X<sub>13</sub> - Collection Size

Perhaps the relationship that has been most frequently tested in reference service evaluation is the association between performance and the size of the library collection. Although size is a quantitative measure, one can argue that larger collections are more likely to provide greater coverage than smaller collections for any given subject.

Although size has frequently been operationalized as number of titles which is an interval measure since each interval of one indicates an equal quantity (i.e., one book). This measure is a poor indicator because not all books are proportionate in the amount of information they contain, nor is the concept of coverage well defined on an interval scale. For example, a library of 100,000 titles has a greater breadth of coverage than a library of 20,000 titles, yet a library of 500,000 titles and a library of 580,000 titles have approximately the same amount of coverage. In each case the difference in number of titles is the same, yet in the former example one library is perceived as being much larger than the other, while in the latter comparison the two libraries are perceived as being approximately the same size. The number of titles required to achieve greater depths of knowledge for any given subject grows exponentially (White 1995: 42-45).

This study will measure levels of collection size in a geometric pattern:

- 7 Over 1,280,000 titles
- 6 1,280,000 titles or less
- 5 640,000 titles or less
- 4 320,000 titles or less
- 3 160,000 titles or less
- 2 80,000 titles or less
- 1 40,000 titles or less

This measure is originally developed for this study. These levels cover the range of library collection sizes in the target population (California State Library 1997). The investigator sorted California public libraries by size and then looked for increases of 5% or more. The levels in this measure loosely describe the natural tiers within the population where these large size increases occur.

#### X<sub>14</sub> - Level of Service

The level of service provided by the librarian is likely to influence the satisfaction of the inquirer and perhaps increase the utility of the information that the inquirer discovers. Wyer described levels of reference service on a scale with three categories: conservative, moderate, and liberal. This scale indicated the amount of effort that would be expended by reference staff in assisting readers. To Wyer, the terms conservative and liberal denote either a thrifty or munificent character of the service (Wyer 1930: 6-10).

In this study, the terms basic, moderate, and extensive are used to preserve the sense of Wyer's descriptors and remove any political connotations. However, it is unlikely that this range of values will show a wide variation within a population of one library type. Generally, the level of reference service provided increases as one compares academic to public to special libraries.

Consequently, the measure used in this study expands the range of values for moderate service to

reveal more variation within the sample. During the orientation, participating librarians will be asked the following:

Choose the statement which best describes the reference service your library provides.

- |   |           |   |
|---|-----------|---|
| 7 | Extensive | Actively performs research. Packages information.   |
| 6 | Extensive | Performs limited research, regularly searches indices (print and online) to locate materials, locates excerpts and passages for readers, packages some information                      |
| 5 | Moderate  | Locates excerpts and passage in sources for reader, guides readers to the stack, recommends particular works, answers ready-reference questions, consults sources, provides instruction |
| 4 | Moderate  | Guides readers to the stacks, recommends particular works, answers ready-reference questions, consults sources, and provides instruction.   |
| 3 | Moderate  | Answers ready-reference questions, consults reference sources, and provides instruction   |
| 2 | Basic     | Directs readers to the catalog, indices, reference sources, and provides some instruction on how to use them efficiently  |
| 1 | Basic     | Directs readers to the catalog, indices, and other reference sources  |

The scores given by each of the librarians for each library will be tabulated and the most common value will be used. This method is preferred to averaging the scores that would reduce the probability of achieving a score of either 1 or 7.

#### X<sub>15</sub> - Management Leadership

The relationship between managerial support and reference performance has rarely been examined even though management has a direct effect on reference staff in terms of working

The relationship between managerial support and reference performance has rarely been examined even though management has a direct effect on reference staff in terms of working conditions, staffing patterns, training, and rewards. Improving the quality of reference service is dependent on understanding this relationship (Aluri 1993: 224-225).

This study will examine the aspect of management leadership in terms of providing staff with a vision of reference service to be provided. This will be measured as follows:

Which statement best describes how expectations of reference service are communicated?

- 7 Comprehensive written reference policy, clear service expectations
- 6 Some written guidelines, general sense of service expectations
- 5 Evolving policy, memos issued in response to problems as they arise
- 4 Oral policy voiced by supervisor in response to problems as they arise
- 3 Service traditions passed among staff, clear understanding
- 2 Service traditions passed among staff, unclear understanding
- 1 No policy of any kind

This scale indicates the method of communication used to share vision with the staff. Increases in value indicate a greater likelihood that expectations are clearly and uniformly understood. The more management explains to staff what should be accomplished, the more leadership is being demonstrated. Birbeck and Whittaker noted in their study that libraries with written reference service policies did not outperform libraries without one, yet their findings are inconclusive because of a small sample size (n=10) and an inadequate performance measure (Birbeck and Whittaker 1987: 64).



## E. Summary of Variables

The three dependent and fifteen independent variables which have been described above are summarized as follows:

### Dependent Variables

#### Reference Service Performance

Y<sub>1</sub> - Accuracy

Y<sub>2</sub> - User Satisfaction

Y<sub>3</sub> - Utility

### First Level Independent Variables

#### Difficulty of the Question

X<sub>1</sub> - Complexity

X<sub>2</sub> - Currency

#### Experience of the User

X<sub>3</sub> - Familiarity with the Library

X<sub>4</sub> - Familiarity with Reference Service

X<sub>5</sub> - Educational Attainment

#### Quality of Communication

X<sub>6</sub> - Approach

X<sub>7</sub> - Interest

X<sub>8</sub> - Understanding

X<sub>9</sub> - Verification

### Second Level Independent Variables

#### Experience of the Librarian

X<sub>10</sub> - Experience

X<sub>11</sub> - Educational Attainment

X<sub>12</sub> - Job Satisfaction

### Third Level Independent Variables

#### Reference Environment

X<sub>13</sub> - Collection size

X<sub>14</sub> - Level of Service

X<sub>15</sub> - Management Leadership

#### 4. CONDUCT OF THE STUDY

Because of the scale and complexity of this study, a pilot study is necessary to uncover potential problems with instruments and measures. The investigator will conduct a small-scale study within the single library to test the logistics of the operational plan. Discussion with librarians and library users will reveal areas of potential confusion or anxiety. Revisions to instruments and operations will be made based on the results of the pilot study.

Letters of invitation to participate in the study will be sent to reference service managers in seventy-six public library jurisdictions in the Los Angeles Basin, comprising Los Angeles, Orange, San Diego, San Bernadino, Riverside, Ventura, and Santa Barbara counties. The invitation will include an example of the statistical report which can be produced from the data gathered during the course of the study in order to help these service managers determine whether the results will be useful or beneficial to them. Twenty libraries will be randomly selected from all those that respond with interest.

The invitation will be extended to central library facilities only. Branch libraries pose unique problems to the investigator because the collections are significantly smaller, the service population includes fewer business, professional, or municipal users, and the small staff would result in every query being answered by two or three librarians. Once the findings of this study provide a better understanding of the variables that contribute to reference performance, an area for future study is the difference in performance between central and branch libraries, both within a single jurisdiction and across jurisdictions.

After the library sites have been selected, staff members from each site will be invited to an orientation. The primary goal of the orientation will be to gain the favorable disposition of the

participants in regards to the study to achieve as high a degree of participation as possible. What is the benefit of having a library manager agree to hold a study when it is the librarians themselves who will be responsible for the quality of the data being collected? The orientation will serve as a “sales pitch” to minimize the amount of animus that staff might have towards the study and maximize the amount of enthusiasm participants have to see the study successfully completed.

To engage the enthusiasm and interest of participating reference librarians, the investigator will instill the librarians with a sense of ownership in the project. The process of generating enthusiasm will be accomplished in three steps. First, librarians will be asked to give authorization for the study to take place. No one can be forced to participate. Of course, unless sufficient numbers of staff at a site agree to participate, then the site will be dropped from the study, and a replacement site will be randomly selected. Second, librarians will be allowed to choose the level of their involvement by electing whether or not to share, in confidence, detailed information about their professional background. Third, librarians at each site will help design the FAQ portion of the instrument that will be unique to their library. The purpose of these actions is to demonstrate concern for participants’ privacy, show respect for the value of their work and knowledge, and to establish a degree of familiarity and trust in order to dismiss any suspicion of the study’s intent. During the orientation session, lunch will be provided as a token of appreciation for the effort the study will require on the part of staff.

Another critical objective of the orientation will be to gather data regarding the service policies at each institution, to record characteristics concerning participating librarians’ experience, and to determine the FAQs for each site. The Library Environment Assessment Instrument (see appendix A), the Librarian Profile Instrument (see appendix B), and the FAQ Survey Instrument (see appendix C) will be utilized to obtain this information.