Evaluation of Reference Service in Public Libraries Using a Hierarchical Linear Model: Applying Multiple Regression Analysis to a Multi-level Research Design

Matthew L. Saxton
Doctoral Student
Department of Library & Information Science
Graduate School of Education & Information Studies
University of California, Los Angeles
4112 W. Franklin Avenue
Fullerton, California 92833
(714) 525-0802; FAX: (714) 525-0810
MSAXTON@UCLA.EDU

December 9, 1997

A Proposal for a Dissertation
in partial fulfillment of the requirements
to be awarded the degree
Doctor of Philosophy

Doctoral Committee:
Dr. Marcia Bates
Dr. John V. Richardson, Chair
Dr. Michael Seltzer
Dr. Virginia Walter

TABLE OF CONTENTS

ABSTRACT

The evaluation of reference service requires a multi-level research design that can account for the effects of intra-class correlation that arise as the result of gathering multiple observations from each library and librarian participating in the study. Few studies have ever tested for these effects which tend to diminish the value of the standard error, misleading the investigator into believing that a relationship between variables has been discovered which does not necessarily exist. Although many previous studies have focused on a single outcome variable, this study will attempt to measure the influence of fifteen predictors on three criterion variables: the accuracy of the information received by the inquirer, the utility of that information, and the inquirer's satisfaction with the reference process. The potential impact of this study is the significant improvement of reference service, improvement of training for reference work, and the improvement of education for reference work. If the objectives of this study are achieved, it will advance the research front in the area of reference service evaluation from an exploratory to an experimental phase.

Budget: $3,525.50

Keywords: Reference Service
Evaluation
Multiple Regression
Hierarchical Linear Models

CHAPTER I
PROBLEM STATEMENT AND THEORETICAL FRAMEWORK

1. PROBLEM STATEMENT

At present, the body of research concerning reference service evaluation is characterized by numerous theoretical and methodological difficulties, which inhibit the discipline from advancing on this research front. These difficulties include:

1) a lack of agreement among researchers as to the definition of reference service, its role in society, and the goals of such service;

2) inconsistency in the operational definitions used from study to study for both independent and outcome variables;

3) the predominate use of univariate statistical procedures that inadequately describe complex social interactions such as a reference interview where the researcher is interested in more than one outcome variable;

4) low sample sizes (a frequent criticism in social science research generally);

5) bias introduced by a lack of random sampling when user samples are self-selected and no effort is made to assess whether or not a representative sample has been captured;

6) little repetition of studies by subsequent investigators in order to confirm or dispute earlier findings;

7) inconsistent reporting of findings that inhibits research synthesis; and

8) a lack of attention given to theory.

As a result, because of this weak theoretical underpinning and some gross deviations from conventional method, published findings are of little value in terms of explaining what factors influence the reference process. After thirty years of research beginning with the two initial studies conducted separately by Charles A. Bunge and Herbert Goldhor in the late 1960's (Bunge 1967; Goldhor 1967), the question of how to evaluate the quality of reference service remains unanswered. No widely accepted method for assessing reference service performance has been established.

Any experienced reference librarian knows that making the library user feel welcome, listening and probing to identify the exact question, and a thorough knowledge of the library resources all contribute towards success in the reference transaction. While these observations are certainly logical, the discipline has not generated the statistical evidence required to confirm or challenge such hypotheses, nor to indicate to what extent these and other factors contribute to a successful outcome. Consequently, despite three decades of research, reference service evaluation remains in an embryonic stage.

Investigators have explored measuring the accuracy of information being provided to users, the satisfaction of users, the job satisfaction of librarians, the efficiency of the service, and various composites of all four. Contributing factors, which have been studied, include both quantitative and qualitative variables, ranging from counting the number of books in the library to asking librarians if they find their work rewarding. At present, this body of research has produced two conclusions: questions are answered accurately about 55% of the time, which is doubted by most in the field, and that library users are highly satisfied and pleased with the service they receive, which is difficult to interpret when given the aforementioned accuracy rating.

The difficulty of interpreting confusing results is compounded by shoddy methodological techniques being employed in many cases. Sample sizes are far too small to defend the generalizability of the findings adequately. Random sampling, the most basic means of eliminating bias, is rarely utilized. Some researchers have used the "regression mill" approach, analyzing numerous quantitative variables without discussing the theory, which would indicate why these variables contribute to reference performance. In anticipating these criticisms, many researchers have described their works as "exploratory" studies or "pilot" studies," thereby cautioning the reader as to the value of the findings and providing an excuse for using small samples. The reader of this literature is left asking, "Where is the actual study? When will the full-scale research be performed?" In many cases, such formal research never took place.

Of additional concern is that fact that the quality of the research can be questioned since it is likely that those primarily responsible for gathering the data, the reference librarians, did not share the enthusiasm of the investigator for obtaining the data. At best, conducting the study was probably viewed as one more task being added to an already overburdening workload. At worst, participants may have feared that performance ratings might be linked to raises and promotions (or lack thereof) and could be used as evidence to build a case for layoffs. Terry Weech and Herbert Goldhor observed that participants in such a study might feel "spied on" (Weech and Goldhor 1982: 306). Unless the investigator can convince the participants that the research project is non-threatening to their position and that the resulting data will ultimately be of some use to them, most reference librarians are likely to view the intent of the study with suspicion.

The lack of a reliable and valid means for empirically assessing the outcome of reference transactions represents a significant knowledge void within the discipline because it obstructs many branches of reference research. This void prevents any meaningful discussion of what methods of professional development and institutional policies are beneficial to the improvement of reference service. In addition, it prevents effective assessment of how well different reference procedures facilitate or hinder service. Furthermore, it impedes the development of sophisticated techniques for collection development evaluation.

However, although earlier efforts have been made to investigate this problem, the traditional use of univariate statistics is unlikely to produce an effective and valid means of assessment because of the complexity of factors that affect the outcome of any reference transaction. Previous studies have usually examined only one outcome variable, and those few cases, which have considered more than one outcome have studied each variable individually.

Perhaps of even greater concern is the lack of attention given to intra-class correlation between libraries and librarians which results from a hierarchical research design. For example, in any study, which involves more than one library, some of the variance in the outcome variables is caused by characteristics unique to the individual libraries. Consequently, the error term is misleadingly small. The investigator is led to believe that the data indicates a relationship exists between dependent and independent variables when in fact the hierarchical effects obscure the true nature of the relationship.

Obtaining a comprehensive understanding of reference transaction outcomes requires an examination of more than one outcome variable. A comprehensive understanding would also require an analysis of the hierarchical effects that arise from the multi-level research design. Only in this manner can the investigator intelligently discuss the amount of variance that is explained by the predictors, the amount of variance that is explained by intra-class correlation, and the amount of variance that is unexplained.

Therefore, a possible solution to the problem of reference service evaluation lies in the application of hierarchical linear modeling (HLM) to describing reference transactions. This method enables the investigator to account for hierarchical effects when more than one observation is gathered from participating libraries and participating librarians.

HLM, which is built upon multiple regression methodology, is not a true multivariate method in the sense that only one dependent value is used in the model (Bernstein 1988: 4; Grimm and Yarnold 1995: 4). This definition of what constitutes multivariate analysis is much narrower than conventional usage twenty years ago where any analysis involving three or more variables could be considered multivariate (Van de Geer 1971: 85-86; Cooley and Lohnes 1971: 3-5). According to this narrow definition, statistical techniques such as multiple regression would be considered univariate procedures.

As has been established above, the outcome of reference service cannot be expressed in a single variable. Such a theory is reflected in the philosophy of the behavioral standards for reference librarians recently adopted by the Reference and User Services Association where the transfer of information is identified as only one facet of the reference process (RASD Ad Hoc Committee on Behavioral Guidelines for Reference and Information Services 1996). Logically, in order to measure the effects of independent variables on more than one outcome variable, it might be desirable to use a multivariate method, such as multivariate analysis of variance (MANOVA) or structural equation modeling (SEM).

Unfortunately, while MANOVA does enable the researcher to test for significant effects on multiple outcomes, it does not provide a means for assessing the magnitude of the effect size whereas multiple regression does. Neither MANOVA nor SEM is able to account for hierarchical effects that obfuscate the strength of association between variables. HLM overcomes both of these inadequacies.

Although HLM may be considered a univariate method, it is possible to create a dependent variable based on more than one dependent measure. Preferably, several outcome variables could be used to create a scale, or could be combined into a single measure through the application of principal component analysis. Thus, the application of this method will bring a high level of sophistication to the field in terms of describing multiple outcomes.

The discovery of a reliable and valid means of measuring reference service is one of the most pressing research needs in library and information science. While this problem is not new, this study will use original methods in attempting to find a solution. The significance of resolving this knowledge void lies in the creation of a stable platform for many other branches of research. The national benefits from this research include:

1) improving curriculum for educating novice librarians,
2) improving staff training programs to enhance skills over the course of a career,
3) analyzing reference queries and user needs as an aid to collection development, and
4) providing information to support decision-making in budget and staff planning.

All four benefits will contribute to a heightened quality of reference service being provided to the service population of all libraries nationwide. The practical applications of this research will take the form of tested training programs for staff, computer software which will generate reports on the distribution of user queries and collection usage, and service policies based on actual user needs.

In addressing the problem of reference service evaluation, this study focuses on reference service in public library settings. In this context, public library users who seek personal reference service are the primary group who will benefit from this project. In the past decade, this group has become increasingly composed of youth, senior citizens, and the economically disenfranchised, as the more affluent members of the community have a growing range of opportunities to access information through the Internet at home, information systems provided by their employers, or fee-based information services. This trend will continue as a larger amount of government information is made available solely through digital media. Thus, the immediate findings of this study will be applicable to reference service being provided to a more disenfranchised population.

2. OBJECTIVES

The goals of this study are to specify a model of reference service that can be used for the purpose of performance evaluation and to test that model at a higher level of methodological sophistication than has been previously recorded or described in the literature. To accomplish these goals, the following three objectives must be attained:

1) To assess the impact of many predictor variables on more than one outcome variable.

2) To assess the effects of intra-class correlation which result from a multi-level research design.

3) To test the reliability of the operational constructs being used for independent and dependent variables.

4) To interpret the apparent relationships between these variables in a manner which is meaningful for the purpose of evaluation.

Each objective is necessary to avoid weaknesses found in earlier studies. The purpose of reference service cannot be captured in a single outcome variable. A hierarchical design is the only way to observe multiple events (i.e. reference transactions) that occur in the same environment (i.e. a library). In addition, testing for reliability is essential when making the argument that a particular construct is a good measure of some phenomenon (just what characteristic is actually being measured by the construct is subject to further discussion). Finally, while the findings of any study may have theoretical value, the results of this study must also have practical value in terms of service planning or decision-making since the ultimate objective of this research is to provide a means for evaluation.

These objectives are original and innovative in that this study will move reference service evaluation beyond an "exploratory" phase of research by conducting a rigorous investigation in accordance with the rules of sophisticated statistical tests. This study is the first effort to apply HLM to the evaluation of reference service.

3. RESEARCH QUESTIONS

In order to obtain all of the data necessary to achieve the objectives enumerated above, the design of this study will be guided by three research questions:

1) What are the independent variables that contribute to high levels of reference performance?

2) What are reliable indicators for both reference performance and the independent variables?

3) How well does the data support the proposed model?

The first question will be addressed in the hypothesis that is used to specify the model. The second question will be addressed through reliability testing of operational constructs used in the study. The third question will be determined by devising a multiple regression equation that predicts a sizable proportion of the variance in the outcome variables (e.g., an equation that predicted forty to fifty percent of the variance would be a notable contribution to the field). If the constructs are judged reliable and the data supports a prediction equation, then a strong argument can be made that the proposed model accounts for the variables that contribute to high performance.

The first question has been asked repeatedly, yet it bears re-examination because of the advantages stemming from HLM methodology. The second question has been asked rarely. Testing for reliability has only been applied to all the variables in a study once (Whitlach 1990b: 67). Unlike construct validity that can only be determined from theoretical premises, construct reliability can be tested statistically for a given data set. The justification for including this question is to obtain evidence to support using a particular construct in anticipation of challenges since no established measures currently exist. The third question has never been asked in regards to a multi-level where the hierarchical effects are examined.

4. HYPOTHESIS

This study will test the hypothesis that reference service performance is determined by the joint effects of the difficulty of the query, the experience of the library user, the quality of the communication between the user and the librarian, the experience of the librarian, and the reference environment. This statement is expressed in a prediction equation as follows:

H1: $y_{kij} = \alpha + \beta x_{kij} + \gamma z_{ki} + \delta w_k + (\upsilon_k + u_{ki} + e_{kij})$

H0: $y_{kij} \neq \alpha + \beta x_{kij} + \gamma z_{ki} + \delta w_k + (\upsilon_k + u_{ki} + e_{kij})$

where y_{kij} are the outcome variables for the *j*th reference transaction completed by the *i*th librarian in the *k*th library, α is the constant, βx_{kij} are the first level independent variables that describe the characteristics unique to the transaction, γz_{ki} are the second level independent variables that describe the characteristics unique to the librarian, δw_{ki} are the third level independent variables that describe characteristics unique to the library, υ_k are the third level fixed effects, u_{ki} are the second level fixed effects, and e_{kij} are the first level random effects. The sum of the fixed and random effects comprises the error term, or unexplained variance, in the equation.

A. Reference Service Outcomes

The first element in this hypothesis, represented in the equation as y_{kij}, is the concept of reference service performance. Reference service, by definition, involves a transfer of information between a librarian and a library user that is initiated in response to a user's query (Wyer 1930: 4; Ranganathan 1940: 53; Hutchins 1944: 10-11; Rothstein 1961: 12). Accordingly, assessing the quality of the performance of that service involves asking 1) how good is the information in terms of answering the query, and 2) how well is the transfer of information accomplished. Thus, the outcome of any reference transaction can be evaluated by three criteria: accuracy, utility, and satisfaction.

If the service is to be of any benefit to the user, then the information being provided by the librarian must be accurate, or in other words, it must truthfully describe reality. Accuracy also implies completeness in terms of addressing all components of the query since a partial answer, though it may be accurate in itself, does not accurately (i.e., exactly and correctly) resolve the user's need.

In addition to accuracy, the information must be in a form that is understandable to the user and can be easily applied to resolving the query. Utility is the degree to which the library user can employ the information they have received through the reference transaction. Utility is a wholly situational characteristic that is dependent on the effective recognition of the user's need by the librarian and the user's abilities to comprehend the material necessary to answer that need.

While accuracy and utility are concerned with the quality of the information being transferred, satisfaction pertains to the process of the transfer. The user must be satisfied that the librarian has provided good service and exhausted all necessary avenues to find an answer to the query. Regardless of how accurate and useful the information may be, if the user believes he or she has been treated badly then he or she is likely to doubt the value of the information. Alternatively, when an accurate and useful answer has not been found, a user who has been treated well is more likely to believe that the librarian has performed well under the circumstances.

B. First Level Predictors

The first level predictors, represented in the equation as βx_{kij}, include variables that measure the difficulty of the query, the experience of the library user, and the quality of the communication between the librarian and library user. These variables will be randomly distributed in the sample.

The difficulty of the query refers to the complexity of the query and the currency of the information being sought. Logically, the more difficult a query is, the less likely an accurate, satisfactory, and useful resolution will be discovered.

The experience of the library user refers to the user's educational level and user's familiarity with the library. The higher the education level, the more likely it becomes that the user is capable of comprehending a wider diversity of materials, both in terms of format and complexity. In addition, a user who is familiar with library services and the manner in which reference transactions are conducted is more likely to establish a better dialogue with the librarian than someone who is unfamiliar.

The quality of the communication refers to how well the librarian and user understand one another and how well the librarian encourages the user to speak freely. Establishing open communication is essential to probing the needs of the user and seeking an answer that is useful. Users are also likely to be more satisfied with a librarian who expresses an active interest in speaking with them

C. Second Level Predictors

The second level predictors, represented in the equation as γz_{ki}, include variables that measure the experience of the librarian. These variables will not be randomly distributed in the sample since each participating librarian will be involved in numerous transactions.

The experience of the librarian refers to the librarian's education, work experience, and service orientation. Logically, a librarian who has studied fundamental library principles, acquired years of experience in practice, and is dedicated to serving the public is far more likely to offer a higher caliber of service than a librarian who has received little instruction, is new to the field, and does not enjoy working with people.

D. Third Level Predictors

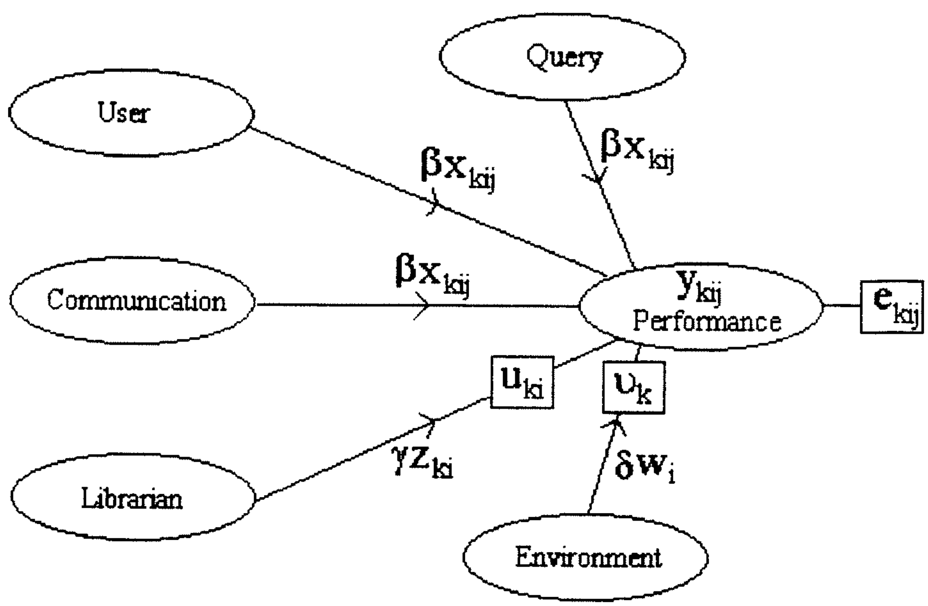
The third level predictors, represented in the equation as δw_k, include variables that measure characteristics of the reference environment. These variables will not be randomly distributed in the sample since numerous transactions will occur within each library participating in the study.

The reference environment refers to the size of the library collection, the service policies of the institution, and the administrative support given to the reference staff. Theoretically, libraries with larger collections are able to provide accurate information on a wider variety of topics a greater proportion of the time than libraries with smaller collections. Well-managed libraries where performance objectives are clearly communicated are also more likely to provide better service.

E. The Reference Service Performance Model

This study will collect data on eighteen parameters that have been discussed above and will be operationally defined as variables in the methodology chapter. The hypothesis states the theoretical relationships between these parameters. Each parameter falls into one of six groups that are graphically described in a model of reference service performance (see Figure 1).

Figure 1. Reference Service Performance Model



5. THEORETICAL FRAMEWORK

To demonstrate that the research questions create a perplexity in the knowledge state of the investigator and that the solution can actually be obtained through means of research (as opposed to deduction), it is necessary to frame the question from a theoretical perspective (Kerlinger 1964: 18-20). A simple means of establishing the theoretical framework is to define the question in terms of the known and unknown (Richardson and Reyes 1995: 238). If the unknown component can be ascertained through observation, then the perplexity can be resolved empirically.

A. The Boundaries of the Study

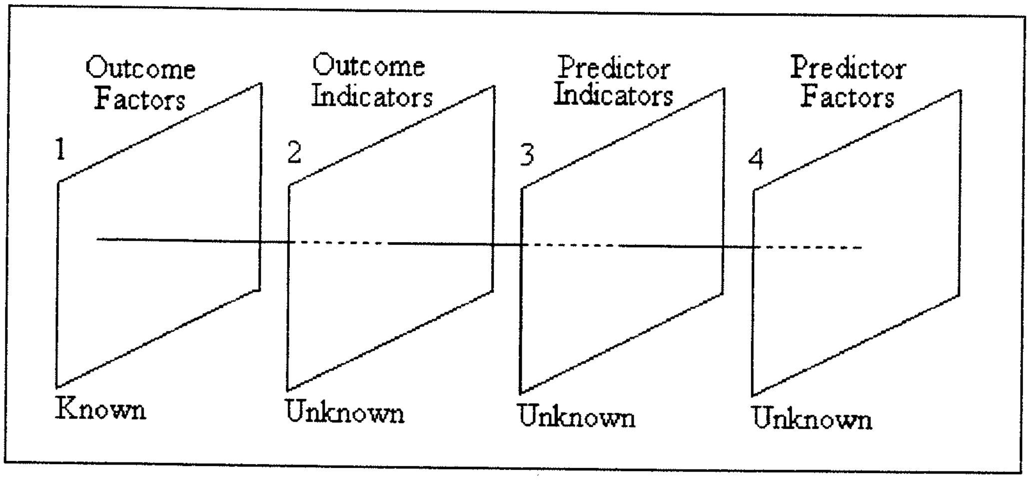
In terms of what is already known, the definition of reference service establishes that high levels of accuracy, utility, and satisfaction are the desirable outcomes for good performance. In terms of the unknown, widely accepted measures for the performance characteristics described above have not been reliably established. In addition neither the factors that contribute to reference service performance nor measures for these factors have been reliably established.

The lack of acceptance or reliability is primarily the result of poor methodology, small sample size, and little repetition from study to study in the research literature. Over thirty-eight variables expressed in 162 constructs have been examined as possible predictors of reference service performance (Saxton 1997: 275-276). Despite this variety, they can all be classified into five groups: variables concerning the reference environment, variables concerning the background and attitudes of the librarian, variables concerning the background and attitudes of the library user, variables concerning the communication between the librarian and the library user, and variables concerning the nature of the reference query. Because of the complex nature of reference transactions, the combined influence of numerous variables from each of the five groups is likely to produce the best model of reference service performance.

This study addresses the unknown through the formulation of a hypothesis as to which variables are the best predictors of reference performance. The hypothesis will be tested through the application of multiple regression analysis in a hierarchical research design in order to determine how well a set of predictors explain the variance in an outcome variable that is some form construct based on three dependent measures. The advantage of this statistical technique is that it enables the researcher to measure the effects of a multiple variables while accounting for the misleading effects of intra-class correlation.

This study can be viewed as an attempt to describe the relationships between four different conceptual levels of knowledge, one known and three unknown (see Figure 2). This study is bounded by a definition of reference service performance outcomes (level 1) and a hypothesis as to what variables can predict performance (level 4). Between these two boundaries, the investigator will determine observable indicators for both outcome and predictor variables (levels 2 and 3). The prediction equation that delineates how the variables influence one another will bridge these conceptual levels.

Figure 2. Theoretical Framework of Four Conceptual Levels



B. Means of Empirical Investigation

Multiple regression is a statistical method where the investigator can assess the ability of multiple independent variables on a dependent variable. Hierarchical Linear Modeling builds on regression analysis by measuring the influence of hierarchical effects in multi-level research designs. In this study, hierarchical linear modeling is used to test the influence of fifteen independent variables in a three-level hierarchical design on three dependent variables. Dependent variables may be examined individually or combined to form a performance scale.

Because of the abstract nature of the concepts being examined in this study, all variables of interest will be measured on ordinal scales. On an ordinal scale, a greater value indicates a greater presence of the characteristic in question. For example, in the case of user satisfaction a score of "5" would indicate greater satisfaction than a score "4." However, ordinal scales differ from interval scales in that the values do not represent proportional quantities of the given characteristic. In other words, a score of "6" would not indicate twice the amount of satisfaction as a score of "3." Ordinal scales are frequently used in social research for measuring human attitudes and opinions that defy physical measurement (Grimm and Yarnold 1995: 5-8).

C. The Unit of Study

Fortunately for researchers, reference service can be analyzed in discrete units known as reference transactions. The *ALA Glossary of Library and Information Science* defines a reference transaction as, "an information contact which involves the use, recommendation, interpretation, or instruction in the use of one or more reference sources, or knowledge of such sources, by a reference staff member" (Young 1983: 189). The information contact occurs between a librarian and a library user, although no mention of the user is made in the above quote. Each transaction begins with a user's query that precipitates an exchange of information between a librarian and a library user and ends when a resolution or a response to the query has been determined. All reference service occurs within the initiation and resolution of the transaction. Therefore, the reference transaction is the subject of this study and not the librarian, the user, or the library, although all three are involved in the transaction.

A useful way to describe a transaction is to classify it according to the type of query that initiates it. Samuel Rothstein reported that reference queries were frequently classified in the literature as one of four types: directional, ready-reference, search (or "research"), and reader's advisory queries. Directional queries involve informing the user about the location of a particular library facility, such as a photocopier or a restroom, or directing them to a specific title that the user requests. Ready-reference queries involve finding a brief fact for the user that is ascertainable from a standard reference source, such as an almanac or encyclopedia. Search queries involve advising the user which sources to consult and teaching the user how to use these sources in order to locate the desired information. Reader's advisory queries involve sharing opinions with the user regarding which books are most suitable for the user's expressed reading interests (Rothstein 1964: 458).

In his popular reference textbook, William A. Katz proposed a different typology which also include directional and ready-reference categories but specifies two levels of search queries: specific search queries and research queries. A specific search query involves finding a single document or section of a document that will answer the question, such as an encyclopedia article. Katz classifies reader's advisory queries in this category. In contrast, a research query requires the user to consult multiple sources in order to develop an answer to the query (Katz 1982: 11-14).

As evidenced by the above examples, the boundary between query types is indistinct. This study will identify queries, and thereby identify transactions, as either ready-reference or research. Any query that requires a short, factual answer from a published source is identified in this study as ready-reference. Research queries are those queries that often require the reader to collect a body of information, often from more than one source, and then evaluate and interpret the information to draw conclusions. The resolution to the query may include creating a search strategy for the reader to follow in addition to directing the reader to specific works.

Directional queries pertaining to library facilities or rules are not true reference queries in that they do not require any knowledge of library practice to answer, nor do they require consulting a source (Hutchins 1944: 16). However, asking for a specific title, which Rothstein classified as directional, will be treated in this study as a ready-reference query since it may involve use of the catalog. Although the answer to reader's advisory queries are usually based more on opinion rather than fact, such queries will also be treated here as ready-reference queries since they generally require a brief answer from the librarian rather than a series of suggestions on how to locate the material.

D. Defining the Query

What is the reference query? Robert S. Taylor defined the query as an entity which undergoes a transformation through four stages, beginning with an undefined, unconscious information need felt by the library user and finally culminating in a formal request by the library user for information. This transformation frequently occurs as a result of the question negotiation process during the reference transaction (Taylor 1962). N.J. Belkin described the unconscious information need as an "anomalous state of knowledge" which discomforts an individual to the point where they are motivated to take action (e.g., driving over to the library) to resolve the anomaly (Belkin, Oddy, and Brooks 1982).

Where does the actual query exist along this continuum from unconscious anomaly to conscious inquiry? At what point in this "maturation" process does the query come into being? Does the query evolve into something new as it passes through each stage, or might one query be answered in many different ways?

For the purposes of this study, the reference query will be defined as the formal request for information negotiated between a library user and a librarian during the reference transaction. The evaluation of service performance will occur in the context of how well the negotiated query is answered. Evaluation in terms of accuracy can only be assessed on the observed query, and not the anomaly that exists in the mind of the library user. If the data which is gathered suggests that accuracy and satisfaction are uncorrelated, this may indicate that librarians are negotiating the query poorly; the user is unsatisfied because the actual query was not addressed, even though the formal query was answered accurately.

An important area for future study is the evaluation the negotiation process itself. A study of this nature would record the initial statements expressed by the library user, note the manner in which the librarian probes the user to determine the actual information need, and compare the final formal request for information to the initial question. Herbert Goldhor has already conducted a pilot study along these lines (Goldhor 1979).

E. Advantages and Disadvantages

The primary advantage that this theoretical structure has which previous studies do not is the recognition of the multi-level nature of the research design. In addition, this structure is predicated upon the formulation of a hypothesis, which has not always been done in the past. Furthermore, the identification of the reference transaction, rather than the library or librarian, as the unit of study is a significant advantage in that the investigator can draw inferences upon the nature of service itself while controlling for the characteristics of the institution or individual which may have a confounding effect on the results.

The main disadvantage, or limitation, of this structure is the reliance on quantitative methods that are unable to capture certain nuances of thought and opinion on the part of users and librarians which the investigator might be able to obtain through survey research. Open-ended responses, which might reveal much information about the reference process, are difficult to quantify and fall outside the boundaries of this study.

This theoretical framework defines the research problem in the context of the current extent of knowledge within the discipline and identifies the knowledge void to be addressed by the study, thereby defining the perplexity being addressed by the investigator. This framework proposes a statistical method for moving from the known to the unknown, and establishes how abstract concepts can be quantified through the use of ordinal variables. This framework conceptually defines the subject of the study as the reference transaction. The query is defined for the purposes of this study. Accordingly, the problem is one that can be resolved through research.

CHAPTER II
LITERATURE REVIEW

This chapter will outline and summarize all past research in the area of reference service evaluation. A discussion of traditional definitions of reference service will lead into a technical description of how those definitions have been operationalized in past studies. This discussion will be followed by an analysis of the two dominant methodologies that have been applied to the study of reference service evaluation in the past three decades: the unobtrusive query-oriented approach and the obtrusive user-oriented approach. This analysis will be followed by a discussion of the application of multivariate analyis to reference service evalautoin, a comment on the impact of obtrusive versus unobtrusive observation on findings, and a criticism of traditional sampling practices. The chapter concludes with a summary of the major causes for reference failure suggested by this body of literature.

This review builds on earlier literature reviews by Samuel Rothstein, Ronald Powell, Kenneth Crews, and John V. Richardson (Rothstein 1964; Powell 1984; Crews 1988; Richardson 1995). Much of the analysis presented here also stems from the present investigator's own research efforts in performing a meta-analysis of reference performance studies (Saxton 1997).

1. DEFINING REFERENCE SERVICE

In 1876, Samuel Swett Green suggested it was important for librarians to provide some assistance to readers in using the library. He believed it was desirable for the library to create an environment which would encourage "personal intercourse between librarian and readers" (Green 1876: 79). The key element in Green's philosophy is the *personal* nature of reference service which represented a departure from the conventional belief that librarians can best serve readers indirectly and impersonally through the acquisition of materials and the creation of finding tools.

Despite initial resistance to Green's ideas at the time, the concept of assisting readers would become accepted as a core function of librarianship. In his *Library Primer*, published in numerous editions near the beginning of this century, John Cotton Dana argued that conversation with readers was essential to determine their wants and make them feel welcome and comfortable in the library. He further wrote that it was necessary for librarians to help train people to use the library in order to stimulate the "inquiring spirit" which was necessary to promote and sustain self-education (Dana 1910: 130-132). To accomplish these goals, Dana instructed librarians to greet persons who enter the library but who might be reticent to approach staff, use tact and patience to "learn at once what the inquirer wishes to know " (i.e. probe), and to show the reader how the information is found (Dana 1910: 58). Although Dana did not explicitly define reference service, he did establish the goals of such service and listed the three techniques described above to accomplish these goals.

In her pioneering *Guide to the Study and Use of Reference Books*, Alice Bertha Kroeger defined reference work as "the assistance given to readers in the use of the resources of the library" (Kroeger 1902: 3). However, this definition was obviously too narrow as she then proceeded to describe in a later edition how a librarian must also search beyond the resources of the local collection as needed in order to address the needs of readers.

> "Other institutions besides libraries are engaged in reference work and are important bureaus of information which the reference librarian should know and refer to when the resources of his library have been exhausted." (Kroger 1908: x)

She further reminds librarians to consider interlibrary loan when assisting "serious readers." Thus, Kroeger expanded the evolving definition of reference service in two ways. First, the reference librarian's responsibility to readers is to guide them to information regardless of where it may be found, establishing referral as a basic principle of reference work. Second, reference service is an activity which is not limited to librarianship. From this latter concept it can be inferred that the practice of reference service must be governed by rules and principles which are applicable across many settings, including differing library settings (Kroeger 1908: ix-xii).

In 1924, William S. Learned wrote that the development of a specialized "intelligence service" is necessary to enable a public library to fulfill the mission of diffusing knowledge throughout a community in order to supply people with information in a form which can "best be utilized by the person in question." Among the qualifications for a person providing such service, he lists "quick intellectual sympathies," which can be interpreted as the ability to grasp another person's thoughts and needs (Learned 1924: 13). Thus, the reason why reference service must be personal is because of individual differences among readers.

> "The chief business of a community library is to produce a general diffusion of knowledge among small, ill-defined, and constantly shifting groups, where each need is peculiar to the individual himself, and must be dealt with separately."(Learned 1924: 27)

Learned believed that "wise personal attention" was necessary to maximize the educational value of a well-ordered library (Learned 1924: 38).

In preparing his own textbook on reference service, James I. Wyer was heavily influenced by Learned and adopted the notion of sympathy in his own definition. He defined reference work as :

> "Sympathetic and informed personal aid in interpreting library collections for study and research...Reference work exists because it is not possible to organize books so mechanically, so perfectly, as to dispense with personal service in their use...It still is, and always will be, imperative to provide human beings as intermediaries between the reader and the right book" (Wyer 1930: 4-5)

Wyer's main contribution is introducing the concepts of interpretation and mediation as the service being provided to the reader, in contrast to aid or direction. The librarian's role is clearly portrayed as something greater than merely pointing the reader to a source. Wyer also wrote that different institutions, as a matter of policy, would provide different levels of reference service, which he labeled as conservative, moderate, or liberal (Wyer 1930: 6-13).

In the late 1930s, both Isadore Gilbert Mudge and Louis Shores helped to expand the definition of reference service by using a functional approach. Mudge wrote that reference work encompassed "*everything* necessary to help the reader in his inquiry [emphasis mine]." This definition goes far beyond the dimension of personal service to include most branches of library work. Mudge judged selection, arrangement, maintenance, and staff supervision to be just as much a part of reference work as instruction and the "constant work in answering individual questions" (Mudge 1936: 1-2). Shores also took a functional approach by stating that the primary purpose of the reference department was to "interpret books to readers" and enumerated five duties of reference staff: answering questions, locating material, performing research, providing instruction, and advising readers (Shores 1937: 2-4). While these five duties all involved personal contact with the reader, in the later *Basic Reference Sources*, Shores expanded his definition of reference work as "everything performed by the reference department" (Shores 1954: 2). Echoing Mudge, this broad definition included supervision and collection appraisal as reference functions.

Shiyali Ramamrita Ranganathan's theory of reference argued that the functional definition espoused by Mudge was too broad, although no evidence exists that he was necessarily familiar with her work. Ranganathan was familiar with Wyer's text, and the following definition is in accordance with the philosophy of Wyer and Learned.

> "Reference service is the process of establishing contact between a reader and his documents in a personal way… It is not possible to do all this for a reader without an intimate understanding of his precise interest at the moment. To get this understanding, there must be an intimate communion between the librarian and the reader… Therefore, reference service is essentially personal service."
> (Ranganathan 1940: 53)

Ranganathan envisioned reference work as a three-step process. He believed that the first step, classification of materials and maintenance of a collection, were preparatory tasks for the second step of reference service, which was initiated by the reader with an inquiry. The last step, which he labeled assimilation, dealt with the ongoing professional development of the librarian who used experiences with readers to prepare for future reference encounters. Service itself is only one phase of a larger process (Ranganathan 1940: 213-214).

In 1944, Margaret Hutchins attempted to draw many threads of reference theory into a coherent form. She stated that reference work encompassed both the personal, direct aid given to individual readers in interpreting library resources *and* the impersonal, indirect activities performed to make information more readily available to readers. She affirmed Kroeger's notions that the librarian should look beyond local resources when helping readers and also that the techniques for performing library service were the same in different institutional settings, regardless of the library's collection or intended service population. She affirmed Wyer's observation that reference service could be provided at different levels according to the policies of the institution.

In 1961, Samuel Rothstein sought to define the dichotomous nature of reference theory more precisely by using two different terms, although in so doing he reversed the conventional usage. Returning to the tradition of Kroeger and Wyer, he defined the personal assistance given to individual readers as *reference work*. This narrow concept is what Ranganathan had referred to as *reference service*. Rothstein then defined the responsibility and organization of a library to prepare for this work as *reference service*. This broad concept is what Mudge, Shores, and Hutchins had referred to as *reference work* (Rothstein 1961: 12).

In 1966, Alan M. Rees also felt the need for greater specificity in describing reference theory. Whether intentionally or unintentionally, he ignored Rothstein's usage and reversed labels again. *Reference service* was the actual provision of assistance, and *reference work* was the library function performed by reference librarians in providing that service. Then he added the term *reference process* to describe the complex interaction between the reader and the librarian as well as the psychological, sociological, and environmental variables that affect them. He explained that the process begins with an inquiry and ends with an answer that is relevant to that inquiry (Rees 1966: 57-58). These terms were later promoted by Frances Neel Cheney in her popular textbook *Fundamental Reference Sources* (Cheney 1971: 5).

This study is concerned with reference service as defined by Ranganathan, wherein the librarian is required to provide personal intermediation between the reader and the information being sought. The reference process, as defined by Rees, is the observable event in which reference service takes place. As such, the reference process, or reference transaction as it now commonly called, is the unit of study that will be examined to draw inferences about the nature of reference service. Reference service evaluation has almost exclusively been concerned with this aspect of reference work, although some recent efforts have gone into exploring the effects of unmediated reference work (Childers 1997: 156-157).

The personal dimension of reference service has been seen as the central defining characteristic from Green up to the present. Why, then, did early researchers ignore the role of the user in designing studies of reference service? Why were reference questions treated as if they existed independently from the context in which they were asked? Beginning in the early years of this century, librarians have been instructed to seek information from external sources when necessary. Why, then, was the importance of evaluating the quality of referrals never discussed? And why were the institutional policies governing the limits of service never controlled as part of a research design?

The literature presents a fairly consistent thread of agreement as to what reference service is. The greatest controversy revolves upon where to draw the line between reference work and other library functions which support reference work. However, many researchers, in seeking to create measures of service performance, departed from conventional definitions and created numerous contradictory and incomplete alternatives which were more easily operationalized. This disagreement resulted in a confused understanding of the role of reference service in society. Information, instead of mediation, was often portrayed as the primary reference product.

2. MEASURING REFERENCE SERVICE

The first investigations into reference service were descriptive rather than experimental in nature. This research was largely concerned with the nature of the types of questions being posed. Edith Guerrier described a project conducted in nine public libraries where every question asked for a week and the source used to answer it was written down. The age of the inquirer and the source used by the librarian was also recorded, although the answer itself was not noted (Guerrier 1936: 529). Florence Van Hoesen undertook an extensive survey of questions being asked at public libraries. Her research lead to the development of a classification of queries which has been used repeatedly in subsequent studies for designing test reference questions (Van Hoesen 1948: 64-67). In neither case were the outcomes of reference transactions a concern to the researcher.

Since the time of these early examinations, models and measures for reference service have consistently grown in complexity. In the 1960s, researchers began with the simplistic model that service quality was equal to the accuracy of the information being provided in the reference transaction. Later, researchers began to express interest in user satisfaction. In addition to these two measures, the present study will add the concept of utility. Thus, instead of replacing accuracy, new variables have been added to explain additional dimensions of reference service outcomes.

Studies in reference service evaluation have historically measured performance according to one of two patterns. The first pattern involved administering reference test questions, both obtrusively and unobtrusively, to librarians and scoring the number of correct answers to obtain an estimate of accuracy. The second pattern consisted of surveying library users, either in person or through a questionnaire, in order to gauge the level of satisfaction with the service. While the first pattern employed proxy inquirers and queries, the second employed genuine inquirers and queries.

A. Tests of Accuracy

The fifty-five percent rule was "established" after a series of reference accuracy studies consistently indicated that just over half of the test questions were answered correctly. The domain of these tests only included ready-reference questions and not any other aspect of personal reference assistance such as directing readers to sources for resolving research queries and providing readers with bibliographic instruction. Although researchers frequently took efforts to ensure that proxy test questions were typical of actual reference queries, no study ever tried to ensure that the proportions of query types on the tests reflected the proportions of query types in practice. To use the findings of these studies to make any general predictions about the quality of reference service is nonsensical.

Of further cause for concern is that these studies are predicated on the belief that reference queries exist as entities outside the context of the inquirers who pose them. Reference service is examined as something performed by the librarian in relation to a question, and not to a reader. Such a concept is a clear contradiction to the reference philosophy of Wyer and Ranganathan.

Why, then, did these tests produce consistent findings if they suffered from the misapplication of theory? In actuality, the test queries in each study were often similar, and in some cases the exact same queries were used. Logically, when similar tests are administered to samples from similar populations, the investigator should expect to see similar results (Douglas 1988: 95-96). The consistent findings suggest a high degree of reliability, but do not prove any degree of validity. Validity can only be argued in light of the theoretical basis of the study.

In 1967, Herbert Goldhor defined reference service as providing answers to fact-type questions and defined the quality of that service as the accuracy of the information being provided. (Goldhor 1967: 29). Looking at reference service as the product of an institution, he hypothesized that accuracy was dependent on the ability of the librarians at that library and the adequacy of the resources in the collection (Goldhor 1967: 10). This conventional view was reflected in *Minimum Standards for Public Library Systems*, published by the Public Library Association the same year, which stated that "quality library service depends upon adequate staff, collection, physical facilities, supplies, and equipment" (Public Library Association 1967). Although Goldhor never wrote it as such, this theory could be expressed as:

Performance = Accuracy = Staff Ability + Library Collection

His research discovered that queries were answered correctly fifty-five percent of the time on average, but he made no attempt to correlate accuracy scores with any independent variables. His theory remained untested.

During the same time that Goldhor was completing his investigation, Charles A. Bunge made the first attempt to link performance with staff ability by examining the effects of education and experience on reference service. He proposed efficiency, defined as the rate of accuracy over the time needed to answer a query, as a better measure of reference performance than accuracy by itself (Bunge 1967: 24-27). He hypothesized that professional training in reference work would enhance performance. His theory could be written as:

$$\text{Performance} = \text{Efficiency} = \frac{\text{Accuracy}}{\text{Time}} = \text{Staff Ability} + \text{Library Collection}$$

$$\text{Staff Ability} = \text{Education} + \text{Experience}$$

The relationship between efficiency and experience was actually expressed negatively. Bunge believed that the difference in performance between trained and untrained staff would decrease as the years of experience increased (Bunge 1967: 30). In other words, lack of education could be offset by experience. Bunge used the Wilcoxon rank order coefficient which was unable to test the cumulative effects of each of the independent variables, but did produce correlation values for individual variables. Bunge's research indicated that education had no significant effect on accuracy, but did result in greater efficiency. Furthermore, the data indicated that the amount of time spent on the reference desk and the amount of time spent in selecting new books (i.e., experience) also had a positive effect on efficiency (Bunge 1967: 61-63). Although efficiency was never widely adopted as an outcome measure, it was used by the Illinois State Library for several reference research projects (Wallace 1983; Wallace 1984; Roy 1985a; Roy 1985b; Drone 1987).

During the year that Bunge's research was published, Terence Crowley explicitly identified five of the six factors that are being used in the present study. He described the reference transaction as a relationship between the inquirer, inquiry, and responder as a dialogue which utilizes a collection of resources prepared by an organization to resolve the inquiry (Crowley 1971: 16-21). In the present study, resources and organization have been integrated in the same factor as Library Environment. Although Crowley identified these five factors, his study would focus on only one independent variable - library expenditures. He hypothesized that those libraries which were able to spend more could build larger collections (resources) and hire and train more competent staff (responders). His theory could read as:

Performance = Accuracy = Library Collection + Staff Ability = Budget

The data indicated that no significant difference in reference performance existed between libraries with high expenditures and libraries with low expenditures (Crowley 1971: 52-53).

These three studies scored accuracy simplistically. Goldhor used a trivariate measure that labeled query resolutions as accurate ($y = 2$), partially accurate ($y = 1$), or inaccurate ($y = 0$). Both Bunge and Crowley both used a bivariate measure that labeled query resolutions as inaccurate ($y = 1$) or inaccurate ($y = 0$). When calculating an accuracy rate, Bunge measured it as a proportion of correct answers out of all queries attempted by the librarian, whereas Crowley calculated an accuracy rate as a proportion of correct answers out of all queries that were posed. Apparently dissatisfied with this rather simplistic measure, Thomas A. Childers introduced a more sophisticated measure that attempted to account for a broader range of outcomes. His system enumerated five different reference outcomes:

1) the answer given is wholly correct,
2) the answer given is wholly correct but uncertainty exists that this is so,
3) the answer given is partially correct,