AN INTERVENTION AND ASSESSMENT TO IMPROVE INFORMATION LITERACY

BY

DAVIDA SCHARF

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in

Communication and Information

Written under the direction of

Daniel O'Connor, Ph.D.

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May 2013

ABSTRACT OF THE DISSERTATION

An Intervention and Assessment to Improve Information Literacy

By DAVIDA SCHARF

Dissertation Director:

Daniel O'Connor, Ph.D.

Purpose: The goal of the study was to test an intervention using a brief essay as an instrument for evaluating higher-order information literacy skills in college students, while accounting for prior conditions such as socioeconomic status and prior academic achievement, and identify other predictors of information literacy through an evaluation of student behavior and attitude.

Design/Methods/Approach: An instructional intervention was evaluated using a brief essay as a pre- and posttest of learning in a course in technical communication. Multiple readers rated essays on five criteria to measure higher-order skills. Interrater reliability and internal consistency of the measures were tested. Analyses of variance and covariance were used to measure academic gains and to partial out the effects of confounding variables. Student behavior was measured by level of activity in the course management system and essay length. Student attitude was measured through a content analysis of their reflective statements. A control group of students who took the same course without the intervention, but who did not take the pretest, also took the posttest.

Findings: 1) The method used for measuring information literacy was found to be reliable and valid. 2) The use of the brief essay as a pre- and posttest showed that the students in the treatment group achieved impressive gains in higher-order skills associated with information literacy. 3) The students in the treatment group significantly outperformed students in the control group with substantive effect sizes explaining results. 4) Socioeconomic status had no significant impact on information literacy. 5) Student use of online instructional materials had no significant impact on information literacy.

Originality/Value: A model of information literacy assessment in higher education was proposed to isolate important classes of variables affecting learning. An experimental design using multivariate methods to account for the multiple influences of variables on information literacy allowed for the determination and partitioning of the influence of each variable and sets of variables. This knowledge allows for efficient and systematic progress to be recorded where less productive variables can be dropped from the model and significant and important variables are kept in the model to increase the amount of variability explained in information literacy outcomes.

ACKNOWLEDGEMENTS

I have been fortunate to have the superb guidance of my advisor and master teacher Dan O'Connor. Sincere thanks to the committee members Claire McInerney, Nick Belkin, and Shanthi Gopalakrishnan for their thoughtful comments and encouragement, as well as to colleagues at NJIT—Heather Dalal, John Coakley, Doris Fleischer, Miriam Ascarelli, Theresa Hunt, MaryBeth Boger, Carol Johnson, James Lipuma, Nina Pardi, Mark Arnowitz, Norbert Elliot, and many others too numerous to name. I am grateful to Irv Katz at the Educational Testing Service who enabled me to participate in a collegial community of information literacy practitioners and scholars. I want to acknowledge the generosity of my many library colleagues all over the country, and especially Denise Koufogiannakis who shared the details of her extensive review with me. I also want to express my deep appreciation for support and insight to friend and colleague Roberta Brody, and fellow doctoral students Paulette Kerr and Jeanette de Richemond. Finally, this work could not have been accomplished without the active engagement of the wonderful students of NJIT.

DEDICATION

Σα βγεις στον πηγαιμό για την Ιθάκη,	When you set out on your journey to Ithaca,
να εύχεσαι νάναι μακρύς ο δρόμος,	pray that the road is long,
γεμάτος περιπέτειες, γεμάτος γνώσεις. . .	full of adventure, full of knowledge. . .
Κωνσταντίνος Π. Καβάφης (1911)	*Constantine P. Cavafy (1911)*

To poet Constantine Cavafy who taught me to appreciate my journey to Ithaka.

To my husband Nikos Robakis, who introduced me to Cavafy, to the importance of asking good questions, and many other wonderful ideas along the way.

To my dear sister Marissa, and to my parents, my children, my grandchildren, and my friends who sustain me in so many ways.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Statement of the Problem

Information literacy has been identified as an important set of competencies for 21st century students and workers. Faculty and librarians in institutions of higher education are seeking ways to assess and improve student competencies in this area. Administrators are being asked by accreditation and government agencies to provide evidence of effective programs for ensuring that graduating students have these competencies. Library administrators and practicing librarians are educators in both a broad and narrow sense, concerned with seeking ways to help their institutions achieve their educational missions by bringing evidence of the value they add to student learning. The Association of College and Research Libraries (ACRL) and others have defined the abilities and some of the behaviors of an information literate person, and these definitions are presented in the Information Literacy Competency Standards for Higher Education. (American Library Association, 2000. See Appendix A.). Librarians have demonstrated that they can teach and evaluate the information handling skills using multiple choice tests, where questions are presented out of the context of an overall research project. For example, limited-response tests have been devised to evaluate whether or not students can find a book in the catalog, use library databases to search for articles, and employ Boolean logic in their search strategies. Yet most of the scholarly literature on information literacy has not gone beyond explanations related to "building block skills," i.e., the retrieval, evaluation, and citation of sources. Little work has been done on Standard Four, which requires methods of assessment of the higher-order thinking skills that are used in contextualized tasks involving secondary research. Although the building block skills are

important, they are not enough. They cannot answer the questions being asked about student outcomes by accrediting agencies because they do not assess higher-level information literacy competencies. Moreover, it would be important to know which factors besides instruction might predict or explain information literacy performance among college students.

This study proposes to address these gaps by developing a model for assessing higher-order information literacy skills and testing several hypotheses implicit in the model. For example, does a prior condition such as socio-economic status and education correlate with student performance on information literacy tasks? Can a diagnostic test be devised that aids learning and provides evidence of teaching effectiveness while accounting for a prior conditions? A good predictor of student achievement, based on a rigorous but practical method for assessing these higher-level skills, is needed. So the first phase of the study involves testing and validating a performance-based instrument used to evaluate the higher-order cognitive competencies associated with contextualized information literacy embodied in ACRL standards three and four (2000). The more complex skills that require integration of other communication skills may be better assessed in context, using authentic assignments. Thus the research design employs a short researched essay as the instrument in a pre- and posttest of learning and achievement following an intervention designed to improve the component and holistic research skills of students. Short diagnostic essays have typically been used for writing placement (O'Neill, Moore & Huot, 2009) and general academic literacy (Bonanno, 2002; Bonanno & Jones, 2007), but not specifically information literacy. Such a test could be used as a diagnostic, as a formative assignment, and as a summative post-test of

individual learning or achievement. As well, it could be locally adapted to serve the purposes of institutional assessment for accountability, as well as program and course assessment all aimed at improving educational outcomes in information literacy and the overall academic performance of students. Once validated, the essay, as a measure of information literacy can also be used to account for the influence of socio-economic and other factors on student learning. The potential significance of developing a reliable and valid method of evaluating information literacy competencies and thus the interventions designed to teach them, using direct assessment of student work at the post-secondary level, cannot be overstated.

1.2 Theoretical Framework

This study spans several theoretical domains. First is the domain of information literacy research, which begins with a definition of information literacy. This study uses the definition presented in the five ACRL Standards for Information Literacy. Although there have been varying definitions proposed by similar groups, they all contain certain foundational elements that define an information literate person. They are the ability to know when information is needed, locate it efficiently, evaluate its quality, and use it to build and communicate new knowledge (ALA, 2000). Theories of educational assessment, especially through the work of Ralph Tyler and his intellectual descendants, Bloom and Krathwohl, were also a significant influence on this study. Tyler (1949) posted four simple questions that tied curriculum, teaching, and assessment together in a tightly bound yet iterative fashion that provided the framework as well as the inspiration for the present study. Tyler asked:

- What educational purposes should the school seek to attain?
- How can learning experiences be selected that are likely to be useful in attaining these objectives?
- How can learning experiences be organized for effective instruction?
- How can the effectiveness of learning experiences be evaluated?

Tyler's ideas laid the foundation for applications in higher education, especially in relation to evaluation of student success for the purpose of institutional evaluation for accreditation, another domain of knowledge that underlies this dissertation. College impact research has produced many models of student behavior, but none include information literacy. One of the theories, Alexander Astin's (1993) Input-Environment-Output (I-E-O) model, which recognizes a role for environmental factors in academic achievement, has been particularly influential on this study. Finally, experimental research methods, frequently used in educational assessment, but rarely in information literacy research, are also a foundation of this study.

1.3 Research Significance

The field of librarianship in general, and the subfield of information literacy instruction in particular, are professions of practice. Thus, evidence-based assessment studies most frequently rely on practitioner observations and user reports. When evaluating the impact of instruction on student essays, few studies in this relatively new area of information literacy assessment research have used experimental designs and multivariate statistical methods that have long been employed in educational assessment. An experimental design using multivariate methods to account for the multiple influences of variables on information literacy allows for the determination and partitioning of the

influence of each variable and sets of variables. This knowledge also allows for efficient and systematic progress to be recorded where less productive variables can be dropped from the model and significant and important variables can be kept in the model to increase the amount of variability explained in information literacy outcomes.

The literature reflects the field's awareness of the importance of assessment, but the rigor common to research studies in other areas of assessment is rare in information literacy research. This study departs from the more typical narratives and case studies in that it employs experimental research designs and statistical data analysis to directly assess student progress in information literacy based on their individual written work.

One likely reason that so few empirical studies have been undertaken is the lack of an accepted conceptual model that can serve as a framework for assessment studies. Lindauer (Gratch-Lindauer, Arp, & Woodard, 2004) must be credited with the first broad conceptualization of information literacy assessment, what she called the three arenas of information literacy assessment, consisting of 1) the learning environment, 2) information literacy program components, and 3) student learning outcomes. She depicted these as three intersecting circles to show that factors beyond the classroom may influence learning outcomes, but its simplicity makes it difficult to use as more than a starting point. Martin pointed this out in his review of a study by Detlor (Detlor, Julien, Willson, Serenko, & Lavallee, 2011) that attempted to use Lindauer's model. Martin's title describes his conclusion succinctly: "Investigation of Factors Affecting Information Literacy Student Learning Outcomes Fails to Undercover Significant Finding" (Martin, 2011). Rather than conceptual models, guidelines and lists of suggestions abound (e.g., Neely, 2006; Radcliff, Jensen, Salem, Burhanna, & Gedeon, 2007). However, in order to

begin to systematically isolate the important variables affecting information literacy, and study their relationships and effect sizes, a comprehensive conceptual framework for assessment studies is needed. Therefore, this study will propose a model for information literacy assessment in higher education. Such a model, if it could be shown to be valid and useful, would be critical to setting the future research agenda for information literacy research. It would provide a framework that could be used to map out the main areas of investigation, breaking the variables into meaningful categories, and showing the interrelationships of those categories. Such a framework could serve researchers as more than just a starting point for study design; it could also function as a structure that could help interpret the effects of the variables that likely affect information literacy learning and achievement. If this framework proves to be useful, it would not only advance conceptual knowledge in our field, but would be helpful to individual investigators as a research planning tool. In addition, it might serve as a tool for communication with institutional and other educational stakeholders about assessing information literacy, which is an important competency in every discipline.

Using the model would be critical to validating it and delineating a replicable process for empirical research in information literacy. Researchers could begin to isolate and test important variables and their relationships, returning to the model to reevaluate the relative importance of many variables. This in turn could enable researchers to break down the cognitive components of the information literacy construct while also identifying and partialing out the effects of influential independent variables, such as prior or environmental conditions, or affective dimensions. Each component of the model and its relationship to other components could be elucidated through research and

would increase our understanding of how and to what extent the variables affect outcomes. Once categories of certain predictor variables and effect sizes are established, it should be possible to learn what helps and what confounds the interpretation of factors affecting the teaching and learning of information competencies. This could help practitioners improve their planning and adjust their teaching focus and methods. Institutions may be better able to understand student outcomes, and to create policies and environments that can mitigate or extend such predictors.

This study will utilize the proposed model by examining a small portion of the model to analyze student performance data based on an authentic assignment. In doing so, this study should be a valuable example to other researchers. The study includes the development and testing of a constructed response assignment, which may provide the first example of an essay designed to be used for diagnostic, instructional, or assessment purposes in information literacy. The study will test an active learning intervention and an assessment rubric under experimental conditions. In summary, using this model would be an important first step in validating the proposed research framework while mapping one of many potential pathways for future research design and analysis.

Finally, the subject of investigation chosen for this study focuses on ACRL Standard Four, an aspect of information literacy that is difficult to study and that has not yet been addressed in much depth. Evaluation of Standard Four requires that the researcher determine whether or not a person is able to use information effectively in order to accomplish a specific purpose. This in turn requires the development of a method and tools for studying contextualized information use. Standard Four may be considered to be at the top of the hierarchy of information literacy skills, because it subsumes all the

other skills under it. Activities related to Standard One (know when information is needed), Standard Two (access it), Standard Three (evaluate it) and Standard Five must precede those of Standard Four (use it) (American Library Association, 2000). Thus, a study that is able to tease out the variables that affect learning outcomes for Standard Four will capture the important conditions of learning of all the information literacy competencies.

CHAPTER 2: LITERATURE REVIEW: EVIDENCE OF INFORMATION LITERACY

2.1 Definition and Theory

The concept of information literacy can be traced to the 19th century and many thoughtful librarians over a 100 year period explored the basic idea of what they framed as user education (Branscomb, 1940; Mathews, 1877; Perkins, 1876; Winsor, 1880; Wriston, 1959). In the last quarter of the 20th century the teaching librarian became more common.

The result was an outpouring of literature on bibliographic instruction (Hardesty, Schmitt, & Tucker, 1986; Kirk, 1975; Tucker, 1980). In 1974, Paul Zurkowski, who was not a librarian, coined the term "information literacy" as part of an analysis of the structure of the information industry. The term reflected important ongoing technological changes and the impact they would have on most aspects of modern life, including education and the role of librarians, but it took several decades for this term to come into general use in the library community.

By 1980, it was recognized that research in user education would depend on the establishment of norms and standards (Werking, 1980). Articulating the competencies that are necessary for a person to be considered information literate became the work of the 1990s. When the concept of information literacy was being discussed by the American Library Association (ALA) (1989), librarians were already exploring the differences between the skill set defined by the older idea (i.e., bibliographic instruction), and the politics and nature of the newer one (Arp, 1990). Information literacy was formally defined and documented by the Association of College and Research Libraries

(ACRL) as a group of competencies that enable a person to 1) know when information is needed; 2) locate it efficiently regardless of its location, format, or medium; 3) evaluate its relevance, authoritativeness, and validity; 4) use it to build new knowledge; and 5) communicate that knowledge (American Library Association, 2000). The ACRL definition has come to be widely accepted in higher education today. It has been endorsed by the American Association for Higher Education and the Council of Independent Colleges (Association of College & Research Libraries, 2004) and has been incorporated into the guidelines of the regional accrediting bodies in higher education (Saunders, 2007). The ACRL standards, however, are not the only definition. ICT (information communication and technology) literacy is used by the Educational Testing Service (ETS), whose work on information literacy began in collaboration with the University of California (International ICT Literacy Panel, 2002). The ETS definition includes an emphasis on technology skills. UCLA's core information literacy competencies differ from those of ETS with its the addition of the ability to navigate a body of knowledge in a discipline (UCLA Library Information Literacy Program Steering Committee, 2005). Outside the United States, Scotland's SCONUL (2004), Christine Bruce's "Seven Faces of Information Literacy" (1997), the Australian and New Zealand information literacy framework: principles, standards and practice (ANZIL) (Bundy, 2004), and UNESCO's primer (Horton, 2008) which is based on the ICT model, are definitions of information literacy with overlapping concepts. O'Connor (2006) provides an analytical criticism of librarians' focus on information literacy as a way to legitimize the profession during a time of change, and Saunders (2010) reviews the lingering debate over the term and its meaning. Despite such criticism, accrediting bodies are adapting the ACRL definition to

different information environments and disciplines (Bonnie Gratch-Lindauer, 2002; also see, e.g., documentation from the Middle States Commission on Higher Education, 2003).

With accrediting bodies expecting outcomes assessment of information literacy, the question of measurement arises. Theories from a variety of disciplines have been used to inform the design and testing of instructional interventions in education. Among the theories used were learning theories of educational psychology based on Vygotsky's (1978) ideas, as well as the work of Bloom and Bruner, have guided studies (Bloom, 1968, 1984; Bloom, Krathwohl, & Masia, 1956; Bruner, 1960). Biggs's Structure of the Observed Learning Outcome (SOLO) Taxonomy goes a step further than Bloom by providing a practical framework for connecting objectives, assignments, and learning assessments (Biggs & Tang, 2011). Mellon showed how Perry's theory of intellectual and ethical development along with Piaget's work on cognitive development could be used as frameworks for user instruction (Mellon & Sass, 1981). Teaching, curriculum, and instructional design theorists (Dick, Carey, & Carey, 2008; Keller, 1987; Tyler, 1949; Wiggins, 1990, 1994) have guided study design as well. Theories of human information behavior from the field of LIS have guided others. Bostick (1992) developed and tested an instrument to measure library anxiety based on the theory first proposed by Mellon (1986). Whitmire (2003) and Gross and Latham (2007) built on this work to study the beliefs and information-seeking behaviors of undergraduates. Macrorie's I-Search (1988) and the guided inquiry model (Kuhlthau, Caspari, & Maniotes, 2007) are student-centered models of active learning that use problem-based learning theory that began with medical education (Albanese & Mitchell, 1993).

In the relatively short time educators have focused on the measurement of information literacy, researchers have developed cognitive and affective measures, using both direct methods (e.g., limited and constructed response tests and performance-based assessments) and indirect techniques (e.g., surveys, focus groups, case studies, and other self-reported and qualitative methods). Two opposing philosophical approaches underlie information literacy research. The first, the scientific method of the ancient Greeks interpreted for social scientists through the positivism of Comte, has come to be represented in the LIS community of researchers largely by the evidenced-based model (Lazarow, 2007). This approach assumes there is an objective reality and that it can be measured, and researchers who follow this approach frequently use limited-response tests and quantitative research methods. These studies must meet standards of reliability and validity. In the second approach, researchers who accept the idea of the social construction of reality seek to account for ways in which a person's understanding is subjective and influenced by interaction with others. Knowledge, even common sense knowledge that is taken for granted, is a result of social interactions, and creates the social world (Berger & Luckmann, 1991). This approach lends itself to qualitative research methods that frequently use data collected through free-response assignments. Some researchers use mixed methods often validated through triangulation. This study focuses on the evidenced-based research model, using a prompted constructed response to provide some of the benefits of both the positivist and constructivist approaches. Prompts and scoring can be tested for reliability, whereas essays provide an authentic contextualized assessment tool that fits well in a classroom environment.

2.2 Higher Education Assessment: Models and Research

"College impact" research, including experimental, methodological, and evaluative studies, form the foundation of this study because it encompasses decades of research on how and why students change during college (Feldman & Newcomb, 1969). These questions lead to an evaluation of the institution's role in student change. In reviews of this literature by Feldman and Newcomb, Tinto (1975), Pascarella and Terenzini (2005), and other studies by important scholars in the field such as Astin (1993), Bean and Eaton, (2000), Kuh, Kinzie, Schuh, and Whitt (2010), and most recently Arum and Roksa (2011), library instruction is never mentioned; indeed, the library itself is only rarely mentioned. However, the questions asked, the methodologies used, and the knowledge gained can guide librarians engaging in research in information literacy. Feldman and Newcomb's far-reaching literature review compared the evidence about ways in which colleges have influenced U.S. college students. The multivariate nature of the study question required a thorough review of studies done between the 1930s and the 1960s. Feldman and Newcomb (1969) found many pitfalls in various methods, but they were still able to articulate nine generalizations that have held up remarkably well since their publication in 1969. They found that college experiences, choice of major, and interaction with peers tend to significantly influence and reinforce students' characteristics. Feldman and Newcomb tackled their research question through close scrutiny of the empirical data and by performing additional analyses of the data on their own. They published the data in a second volume to encourage examination of the many decisions they had made in order to compare the data. Within-institutional studies often face the same challenges in comparing multiple studies across the curriculum.

The next generation of college impact research began by reviewing previous work in order to gain a better understanding of why students fail to finish college. Unlike Feldman and Newcomb, who asked about all college students and how they changed or what they had gained, Tinto's (1975) work was a search for why students fail. Based on Durkheim's explanatory model of suicide, Tinto theorized that family background, individual attributes, abilities, and prior education lead students to certain levels of commitment to personal goals and commitment to the institution. Much as Feldman and Newcomb had observed, these prior inputs are influenced and often reinforced by the academic and social systems in college. In Tinto's model, students need social and academic integration to persist in college.

Tinto began an era of research that sought to identify relationships between the environmental factors that influenced student engagement and that in turn affected student retention, especially in the first year. Tinto's (1975) "Conceptual Schema for Dropout from College" and later Astin's I-E-O model (inputs, environment, outcomes) acknowledge the preconditions and institutional factors influencing student success. Astin proposed a college impact model to assess the influence of college on student outcomes that focused on the relationships among the variables over which educators could exercise control. His main concern was student learning. He saw its purpose as one of "talent development." His theory of student involvement places a high degree of importance on motivation and led to the conclusion that one way to assess teaching might be to measure the degree of motivation fostered by the instructor. Reason, after reexamining persistence research in 2009, concurred with Tinto and others that student engagement and persistence are significantly correlated. Pascarella (2005) added

structural and organizational characteristics of institutions as an important category of variables that influence student success in college. The theories of student change have been based on the psychosocial theories of Erikson and the cognitive theories of Piaget.

Student learning is also affected by choices in curriculum and instruction, as shown by Tyler's (1942) landmark study on curriculum, the Eight-Year Study, conducted in the 1930s. He conducted a longitudinal analysis of 30 schools and the careers of their students in order to evaluate the effects of progressive programs compared to schools using conventional curricula. He used matched pairs of students and measured grades, extracurricular participation, dropout rates, intellectual curiosity, and resourcefulness, concluding that, according to every measure, students following the experimental curriculum did as well or better than the control group. "Types of instruction," which is closely related to curriculum, has been another research thread. Braxton, Milem, and Sullivan (2000) studied the relationship of active learning to student persistence and demonstrated that active learning had a significant impact on social integration and intent to return. This is somewhat surprising because one usually thinks of social conditions affecting learning rather than learning conditions affecting the social aspects of student life. Rather than demonstrating the direct impact of a teaching technique on learning, Braxton demonstrated an indirect effect of a teaching technique on persistence.

In addition, many researchers have sought to elucidate the relationship between socioeconomic status (SES) and academic achievement. White (1982) reviewed these studies and Sirin (2005) used White's research design in a subsequent review. In both cases, medium to high correlations were found between SES and academic achievement. Questioning how much of the relationship between admissions tests and academic

performance in college is explained by SES, researchers at the University of Minnesota partialled out the effect of SES on each. They showed that SES influences test scores, and test scores influence grades, but that the relationships are not an artifact of the common influence of SES (Sackett, Kuncel, Arneson, Cooper, & Waters, 2009). The implications of these findings make it imperative that not only the effects of SES be taken into account, but also the way in which SES data is collected and analyzed.

Best practices methodology, a qualitative approach, was used on a large scale research project, Documenting Effective Educational Practice (DEEP), which was designed to study institutional effectiveness. Kuh, Kinzie, Schuh, and Whitt (2010) developed a set of criteria and a methodology for selection of 20 institutions with high performance in student engagement and graduation rates while taking into account student and institutional demographic factors. They sought to identify a set of characteristics common to these successful institutions. The institutions' libraries were mentioned only in passing and it is not clear whether that was because its various roles were not considered or because none of the libraries had a recognizable impact on student engagement or graduation.

2.2.1 Evaluation of Academic Achievement

College impact research spawned a literature on academic achievement and program evaluation that bridged research and practice, a space familiar to scholars of information literacy. Theoretical and practical professional guidance such as that of Palomba and Banta (1999), Weiss (1998), Suskie (2009), and Middaugh (2010) provide tools for evaluators and researchers, who must choose among purposes, methods, and processes when designing and carrying out meaningful studies. Guided by accreditation

requirements, these practitioners advise that only multiple measures can adequately document improvement in a student's cognitive development and the institution's role in achieving them. They suggest using both standardized and locally developed tests as well as more comprehensive demonstrations of student work such as that of portfolios and capstone projects. Likewise, the literature on methods (Krathwohl, 1998; Trochim, 2001) and measurement (Pedhazur & Schmelkin, 1991; Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007; Shavelson, 2009) provides many models of experimental design and analysis for evaluative studies in higher education. Shadish, Cook, and Leviton (1991) organized the theories of practice into three parts: methodologies, links to policy, and the integration of the first two. They credited Scriven's general theory of evaluation with laying out an overall approach still taken by higher education evaluators today. It consists of three elements: selecting criteria, setting standards, and assessing performance.

Although traditional experimental designs that use random assignment and control groups would be preferable ways to reduce complexity in analysis and threats to validity, these methods are only rarely feasible for instructional librarians. The existing conditions under which students enroll in particular courses, select particular professors, or programs of study are factors that are virtually impossible to control and difficult to control for. Thus, quasi-experimental designs have played a prominent role in education research. Because a course of study in college lasts several years, it is necessary to conduct longitudinal analyses, although changes in curriculum and admissions criteria may confound interpretation.

Qualitative methods may be used, such as focus groups and case studies, which are particularly appropriate in this context, and can provide insight and support in

advance of a research design. In my experience, and suggested in analyses of the politics of program implementation and evaluation (Weiss, 1998), an important impediment to good assessment is librarians' and faculty members' lack of time and conflicting priorities. Thus, Shadish's third element of evaluation theory, methods of integrating methods and policy, must also be addressed. This can be done through collaborative work, as well as at the levels within the institution where strategic goals and objectives are set for programs and operating units that provide support such as the library, labs, and computing infrastructure. This process in itself can become an object of evaluation if such objectives have been set. For example, if a university-wide objective requires each program to include information literacy as a learning outcome, then a systematic qualitative evaluation of assessment reports in each of the designated courses where information literacy is taught could provide evidence of process.

Methods of assessing student learning, achievement, and institutional effectiveness have been summarized by Pascarella and Terenzini (2005), Kuh, Kinzie, Schuh, and Whitt (2010), and Shavelson (2009). To obtain meaningful results with so many variables, these college impact researchers have employed rigorous experimental designs and used multivariate statistical methods for data analysis. Although data are most often collected on individual students, the data may also be aggregated to a cohort of students representing a particular trait, class, program, or category of institution. As the purpose of the analysis changes, so does the unit of analysis. Aggregated data can mask and or unmask important information. A deeper analysis of the college impact literature can be informative, but is too voluminous for this short review. Perhaps it is also too expensive for busy librarians because according to Jackson (2007), Ondrusek (2008)and Saunders

(2009) librarians still have a lot to learn about theory and practice from developmental, educational and behavioral psychologists, educators, instructional designers and assessment professionals. Or perhaps the chaotic nature of information literacy research is not due to a lack of models or time, but to the preliminary nature of the research that has been conducted thus far, and to the diversity of theories and models already found to be useful.

2.2.2 Review of the Library Impact Literature

Although empirical studies of library impact have been generally absent from the literature on student success, there are some exceptions. In the 19th century, some librarians saw the importance and potential impact of "use instruction" (Mathews, 1877; Perkins, 1876; Winsor, 1880). However, the emphasis on inputs continued for decades. The first studies, like most evaluative studies in higher education, were largely characterized by the collection of data such as the number of volumes in the library, library expenditures, or number of reference questions answered. This approach continues right up to the present day with academic libraries still reporting data on inputs such as collection, service transactions, staff and expenditure size to the U.S. Department of Education (National Center for Education Statistics).

A few studies found positive relationships between library use and student persistence. Branscomb (1940) and Barkey (1965) were notable for their comparisons of the number of books a student borrowed and his or her scholastic standing. In 1963–1964 a data analysis comparing student library workers' retention data with that of the general undergraduate population was conducted at a California state college. A significant correlation was found between library use and persistence among freshmen in arts and

agriculture, but not in science and engineering (Kramer & Kramer, 1968). A study by Mallinckrodt and Sedlacek (1987) surveyed a stratified random sample of second semester freshmen. The authors found that the use of campus facilities in general and of the library specifically were positively related to student retention, especially for African Americans. The survey included separate questions on the use of the library both as a place to study and for its use for research. A similar result was found by Rushing and Poole (2002) at Loyola University in New Orleans. They hypothesized that working in the library increased both library knowledge and student involvement and they found a significantly higher graduation rate for student library workers in general, and an even higher rate for minority students.

In Boyer's study of the college experience (1987), one chapter discussed libraries and echoed Farber's belief that libraries should focus on user instruction, not materials (Farber, 1974). Others took up the theme of the library's influence on student engagement and success using national survey data. They employed statistical techniques such as multiple regression in an attempt to isolate the effects of the variables of interest. Kuh and Gonyea (2003), for example, used regression models to analyze data from hundreds of thousands of student responses to the College Student Experiences Questionnaire (CSEQ) across the United States over several decades. Library use was measured by survey questions of self-reported usage of library content. The authors did not find that library use made an independent contribution to student outcomes. In this study, information literacy was one of the outcome variables and was measured by a separate set of questions regarding students' estimation of their own progress in information handling, critical thinking, and the use of technology. The authors identified

several possible causes of their failure to find a direct contribution of library instruction to students' perceived gains in information literacy: 1) validity of their scale as a measure, 2) the lack of baseline measures of information literacy for comparison, and 3) the high number of other experiences beyond the library instruction that can contribute to information literacy and critical thinking skills. All three of these limitations are persistent problems for researchers in this field and ones that are addressed in this study. As they pointed out, without baseline and longitudinal data, one cannot quantify the gains achieved during college. Using national survey data is one way to benchmark outcomes across institutions, an activity critical to accountable stakeholders. It has been suggested that the National Survey of Student Engagement (NSSE) could also be a useful instrument (Mark & Boruff-Jones, 2003). An ACRL committee did develop a set of questions for inclusion in NSSE. They were piloted, but they have not become a standard part of the survey (Gratch-Lindauer, 2007). Several librarians have sustained the effort to collect and analyze longitudinal data within institutions (Breivik, 1974; Hardesty, Lovrich, & Mannon, 1982; Selegean, Thomas, & Richman, 1983).

The concept of return on investment (ROI) has been frequently taken up by supporters of special libraries and public libraries, where accountability has long been important. They used criteria and measures different from those used in the valuation of libraries in higher education, that were frequently based on the value of providing collections, which is backwards-focused, rather than looking forward to redefining the academic library in terms of function (Bailin & Grafstein, 2005; Kelly, 1995). Likewise, the work done on LibQual, a tool for gathering and analyzing academic library data for assessment, focuses primarily on the collection and access side of the library (Cook &

Maciel, 2010; LibQual+). The literature is large and tangential to this discussion. Nevertheless, their employment of experimental and statistical methods is instructive. Hamrick, Schuh, and Shelley (2004) found that "library expenditures provide a very robust and statistically significant explanation of graduation rates" (p. 12). Using multiple regression, they were able to isolate the impact of instruction, but did not separate library instruction. Indeed Hamrick's "library" variables still assumed the "library as collection" model as the predominant model. In another study of ROI and retention, Mezick (2007) compared various categories of library expenditures and the number of professional library staff to student persistence and found some significant correlations. Both ROI studies employed national statistical data to identify a relationship between persistence and libraries in U.S. colleges.

The role of the library collection, library services, and information literacy instruction in helping an institution of higher education achieve its educational goals is an area where further study could make a significant contribution. The library has often been referred to as the heart of the university, but the evidence of its impact has been largely absent from the mainstream research in higher education, although student information literacy is one measure of institutional achievement. One way that academic libraries support information literacy goals is by teaching and building collegial relationships with students and faculty that promote engagement and learning. The effect of librarian-faculty collaboration is another thread in the academic library literature, especially in relation to college composition where it is discussed as having a positive effect on information literacy (J. Elmborg & S. Hook, 2005; Knapp, 1966; Leckie, 1996; Norgaard, 2003; Samson, 2010).

2.3 Measurement of Cognitive Aspects

2.3.1 Limited-Response Tests

The initial focus of information literacy assessment was on the cognitive aspects of student learning as evaluated by the use of limited-response tests. These tests consist of multiple-choice, true-false, and matching questions, which have specific answers and can be scored objectively and reliably. The most relevant of these among the national tests of general education are the Proficiency Profile, and the Collegiate Assessment of Academic Proficiency (CAAP). The Proficiency Profile from the Educational Testing Service (ETS) measures general education and critical thinking (Young, 2007) using solely multiple-choice items. CAAP is a nationally normed test used to measure general education program outcomes, typically in the sophomore year, and consists of multiple-choice and constructed response items. The Collegiate Learning Assessment (CLA), a more recent development, uses student writing to measure the impact of four years of college, (Benjamin et al., 2009).

The field of psychometrics addresses issues related to the construction and testing of reliable and valid instruments for measuring psychological traits or abilities, but only a few librarians are familiar with psychometric theories and methods. Academic librarians often use locally developed limited-response tests to measure a student's command of the material in a workshop or course. These tests are quickly developed, rarely pre-tested, and most often presented as case studies or practice guides if they do reach the library literature (Oakleaf, 2008). Some standardized limited-response tests have been developed by librarians based on testing theory, and they have evaluated and reported their reliability and validity. Of these, the test most frequently mentioned is the Standardized

Assessment of Information Literacy Skills (SAILS), designed for large-scale administration to undergraduates. One adaption of SAILS for graduate education students was the Beile Test of Information Literacy for Education (B-TILED) (Beile-O'Neil, 2005). SAILS is based on item-response theory and was developed at Kent State (O'Connor, Radcliff, & Gedeon, 2001). Similar tests were developed for beginning undergraduates at James Madison University (Cameron, Wise, & Lottridge, 2007), University of Maryland (Mulherrin & Abdul-Hamid, 2010), City University of New York (Ondrusek, Dent, & Bonadie, 2005), and South Dakota State University's Information Literacy Exam (SDILE) (Leibiger and Schweinle, 2009). The work that has gone into SAILS and other tests has provided librarians with tools for measurement and assessment. They can be useful as one among several measures, especially because they may help determine the validity of new information literacy instruments.

Two other tests have gone beyond the usual types of questions used on the limited-response tests. The first, iSkills, is a unique computer-based adaptive test that lies somewhere between limited and constructed response tests, and was designed to assess information literacy skills developed at ETS. It is based on the information-communication-technology (ICT) literacy construct as defined by ETS to contain the following categories: define, access, manage, integrate, evaluate, create, and communicate. Cognitive skills, traditional literacy, and technology skills are critical components of the construct (Katz, 2007). Test-takers must respond to complex test items depicting scenarios that require what seem like constructed responses, but are actually restricted to a limited set of choices. Katz theorized that such a performance-based test would be a more valid measure, but to ensure reliability the responses need to have a

bounded range. The second test is the Research Readiness Self Assessment (RRSA), a limited-response test. Unlike iSkills, it is a locally developed, performance-based multiple choice test, and some questions require the students to compare and evaluate the quality of sources. Created and tested by a psychology professor concerned about her students' lack of research skills, it was validated and used as a diagnostic early in the term with positive results (Ivanitskaya, Laus, & Casey, 2005). It is also unique in its incorporation of three types of questions: multiple choice questions, skill-based problems, and attitudinal measures.

Ivanitskaya, in a subsequent study using the RRSA, found the effect of the pretest to be positive on both skill level and attitude among college students, although the sample size was small (n=32) (Ivanitskaya, DuFord, Craig, & Casey, 2008; Zehner, 2010). Zehner's (2010) doctoral study used the RRSA with a larger sample of high school students (n=170) and found similar results. Indeed, high school students pretested with the RRSA did better than a control group of first-year college students in several areas.

Why not use these standardized tests to measure information literacy? The construct of information literacy itself may still need validation. Surprisingly, a study comparing iSkills to a reliable and valid locally developed constructed response assessment (both tests purported to measure the ACRL information literacy construct) failed to find significant correlations between the two (Katz et al., 2008). A more recent study by Beile, Dziuban, Katz, and Salem (2010), designed to investigate this problem compared scores on iSkills and SAILS, found moderate correlations. A correlation of .56 indicates the tests overlap significantly, but perhaps not as much as one might hypothesize. The authors concluded that the specific abilities identified by ACRL may

not be distinguishable by such a test. As general tests of ability, it has not been clearly shown that such tests can isolate the effect of specific instruction intended to enhance students' overall information literacy. Moreover, these tests are not able to measure the most complex and ultimate goal of information literacy: the ability to use information in the creation of original work.

Another significant problem is that of student motivation. It is difficult to create incentives for students not only to take these tests, but to try their best, as Hawthorne (2008) and Banta (2008) both mention in their criticism of the standardized testing approach used by the Voluntary System of Accountability (VSA) and others. Furthermore, neither the information literacy-specific, nor the general education limited-response tests measure ACRL Standard Four. Although these tests have been shown to be reliable and valid for the skills they test, their main problem is their inability to measure higher-order skills. If information literacy is a holistic competency, we might ask if there is a critical thinking component that transcends the standards and performance indicators described by ACRL. Weiner (2011) contends that there is substantial overlap between information literacy and critical thinking and that information literacy is involved in all of the cognitive functions suggested in Bloom's taxonomy. Yet, few measures, even those developed by librarians, assess the ACRL standard most closely associated with critical thinking, ACRL Standard Four, i.e., the utilization and integration of knowledge gained from information sources for a specific purpose (ACRL, 2000).

2.3.2 Constructed Response Tests

Unlike limited-response tests, where a test-taker chooses an answer, constructed-response tests require the test-taker to produce an answer. A typical limited-response

question in information literacy might ask a student to select which among several library databases would be most relevant for answering a question. A constructed-response question might ask students to write an essay on a particular topic using sources of their own choosing. Although the limited-response test can test the student's command of the facts and her or his ability to discriminate among several choices, the constructed-response test requires an original and more complex response. The demonstration of ACRL Standard Four requires original performance, thus a constructed response. Although well-suited to testing complex skills, constructed-response tests are also more time-consuming to answer and score. Limited-response tests can be more reliably scored, but are often challenged as lacking validity. Constructed-response tests are more easily validated, but more difficult to score reliably. Although there are some computer programs that can score writing abilities, the technology is most used in high volume testing. It is a mature technology, but it still engenders controversy (McCurry, 2010). When conducting human scoring, maintaining consistency can be a task that often requires resources not always available to testers, even for relatively small samples. In any case, although computer programs exist for scoring writing, none have been developed for scoring information literacy abilities in authentic assignments.

To address the problem of reliability, rubrics are used to maintain consistency among scorers, enabling the testers to code a qualitative assessment of an authentic student work product as a quantitative score. Multiple readings of a single text by trained scorers familiar with the subject matter are typically employed to ensure inter-rater reliability (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). There are two

types of scoring rubrics used with constructed-response items—analytic and holistic rubrics. “Analytic rubrics” specificy the elements that must appear in the response, whereas “holistic rubrics” describe likely characteristics (and provide examples). A student may possess the component skills, yet the holistic view may reveal a less competent performance than an evaluation of the individual parts would indicate. Scoring rubrics sometimes contain both types of guidelines for scoring different elements of the same test item (Moskal, 2000). In addition to reliability issues, and the time-consuming nature of the scoring process, another limitation with assessing authentic student work is validity (Wiggins, 1994, 1998). The literature on using rubrics for educational assessment is large, yet a recent review article identified many areas where further research is still needed (Reddy & Andrade, 2010). Among the research gaps identified is the paucity of robust studies on the reliability and validity of rubric use, and the lack of focus on learning and rubric use in postsecondary education.

The use of rubrics as a tool for information literacy assessment is relatively recent. Such assessment includes the evaluation of research papers, bibliographies, and portfolios of student work (Oakleaf, 2010; Rosenblatt, 2010; Scharf, Elliot, Huey, Briller, & Joshi, 2007; Walsh, 2009). Rubrics have increasingly been used to control interrater reliability in conjunction with student research products (Choinski & Emanuel, 2006; Choinski, Mark, & Murphey, 2003; Daniels, 2010; Emmons & Martin, 2002; Green & Bowser, 2006; Knight, 2006; Kohl & Wilson, 1986; Oakleaf, 2006; Scharf et al., 2007; Snavely & Wright, 2003; Sonley, Turner, Myer, & Cotton, 2007; Van Helvoort, 2010). The Institute for Museum and Library Services (IMLS) has funded a cross-institutional

research project to address the use of rubrics in information literacy. Results thus far have addressed process, focusing primarily on ACRL standards 2 and 5 (IMLS, 2012).

If information literacy is indeed a liberal art, as Shapiro and Hughes (1996) proposed, then measuring performance will be fraught with all the problems encountered when assessing general education skills, which are difficult to separate from cultural and disciplinary frameworks and knowledge. One method that attempts an authentic integrated assessment of academic literacy is Measuring the Academic Skills of University Students (MASUS)(Bonanno & Jones, 2007). It was funded by the Australian government and uses a short essay that tests some aspects of information literacy. It has been used at the University of Sydney in Australia since 1990 and has been found to be both valid and reliable (Bonanno & Jones, 2007; Erling & Richardson, 2010). The first of the four major criteria assessed by MASUS is "information retrieval and processing of verbal, visual and numerical data—is this accurate and appropriate to the task?" The second considers "structure and development of text—is this clear and generically appropriate to the task and its context?" The remaining two of the four criteria refer to the structure, style, and form of the essay. This method uses a constructed response to measure some complex skills associated with writing and research, but fails to isolate ACRL Standard Four.

2.4 Measurement of Affective Aspects

Although the major focus of information literacy assessment thus far has been on cognitive achievements, the affective aspects on information literacy learning have not been entirely ignored. The affective dimension in information studies was recognized by early educational theorists such as Bloom (1956) and Bruner (1960), as well as by

theorists in information studies, such as Wilson (1981) and Dervin (1986), on which Kuhlthau's work (1988) was built. Neely (2000) found significant correlations of sociological and psychological factors such as exposure, experience, and attitude with information literacy skills, measured through student self-reports. Nahl (1996; Nahl & Bilal, 2007) and Farmer (2007) have studied the relationship between information and emotion, particularly the role of self-efficacy (as defined by Bandura, 1997) and self-regulation (as developed by Schunk and Zimmerman; see Schunk, 1995; Schunk & Zimmerman, 1994; and Zimmerman, Schunk, & Martin, 2004). These concepts also underlie some of the studies that used survey instruments designed to measure student self-efficacy in relation to information literacy competencies (Kurbanoglu, Akkoyunlu, & Umay, 2006; Pinto, 2009). Zimmerman and Kitsantas (2002) tested the theory of self-regulation in the context of a writing assignment and found that social feedback positively affected learning outcomes. Studies such as those of Jacobson and Xu (2002) focused on the aspects of instruction that influence motivation. Students' reflective journals have been studied for examples of engagement and satisfaction (see, e.g., McGuinness, 2007). Linking cognitive and affective behaviors by comparing scores on validated instruments measuring critical thinking and library anxiety confirmed an inverse correlation between the two constructs (Kwon, 2008; Kwon, Onwuegbuzie, & Alexander, 2007). Overall satisfaction with library services has been measured on a large scale using national surveys of student satisfaction and engagement such as the College Students Experience Questionnaire (CSEQ) (Kuh & Gonyea, 2003) and the National Survey of Student Engagement (NSSE) (Mark & Boruff-Jones, 2003). Additional

questions about student engagement with a library's information literacy efforts were piloted in 2006 (Gratch-Lindauer, 2007), but have not been added to the NSSE survey.

2.5 Writing Assessment Models

Writing and research processes have strong similarities as noted by scholars in both fields (Bowles-Terry, Davis, & Holliday, 2010; D'Angelo & Maid, 2004; J. Elmborg & S. Hook, 2005; Norgaard, 2003). The struggle of librarians to define information literacy in the 1990s resembled the struggle of college composition teachers to define effective writing several decades earlier (Smagorinsky, 2006). The iterative nature of writing and research is understood by composition teachers and teaching librarians, but is difficult to communicate to learners in a framework in which product rather than process has been traditionally emphasized. Thus, both disciplines have moved towards a process model of teaching writing (Emig, 1971; Flower & Hayes, 1979) and research (Andretta, 2011; Kuhlthau, 1988). Both fields have been influenced by the theory that meaning and knowledge are built through scaffolded learning, using a series of guided activities, practice, and reflection. Both fields recognize the importance that affective aspects play in the learning process in terms of motivation and self-efficacy in writing (Hidi & Boscolo, 2008; Pajares & Valiante, 2008) and research (Kurbanoglu et al., 2006; Pinto, 2009). Both have, at times, addressed the acquisition of component skills in citing sources in the context of dealing with student plagiarism (Howard, Serviss, & Rodrigue, 2010; P. A. Jackson, 2006; Ritter, 2005). Indeed, Elmborg and Hook (2005) suggest that the two processes are so intertwined that separating them risks compromising our understanding of how students work and how best to facilitate writing with sources. Moreover, scholars and practitioners in both fields embrace the notion of embedding

learning opportunities across the curriculum (Bruce, 2004; Elmborg, 2003; National Council of Teachers of English NCTE, 1983). Both see that critical thinking skills may be generalizable across disciplines, but that contextualized disciplinary thinking is essential in both skill areas (Bazerman & Russell, 1994; Grafstein, 2002). However, differences in purpose and scope were noted by Fister (1993) in her examination of the rhetorical dimensions of research. Hers is one of the earliest pleas to librarians to consider how a research paper assignment can enable the integration of bibliographic and composition instruction.

Given so many similarities, the history of writing assessment may provide a good model for information literacy assessment so far, and an indication of where it may be headed. Yancey (1999) characterized three overlapping waves of writing assessment—objective testing from 1950 to 1970, holistically scored essays from 1970 to 1986, and portfolio and programmatic assessment, from 1986 onward, with one method never completely displacing the previous ones. Her description of the rationale for the progression from one phase to another may be instructive for teaching librarians. Prior to the 1970s, writing instruction was prescriptive, with the use of books on grammar and syntax represented by high school English textbooks (e.g., Warriner, 1957). To test such instruction, the writing assessment utilized indirect, limited-response assessments with questions about grammar and syntax that led to automated scoring. Although information literacy has not been encoded into an iconic textbook in this way, the existing texts and online tutorials do focus largely on ACRL Standards Two, Three, and Five, the "finding," "evaluating," and "citing" skills, which are more easily made prescriptive than Standards One and Four. For example, see the sections on research in widely used composition

textbooks published by St. Martin's (e.g. Hacker, 2011) or Longman (e.g. Ruszkiewicz, 2011), or even those authored by librarians that focus solely on research skills. A good example can be seen in the table of contents of a librarian's textbook now in its fifth edition, *The College Student's Research Companion* (Quaratiello & Devine, 2011). Tutorials featured on the ALA peer-reviewed database, PRIMO (Association of College and Research Libraries, 2012) are likewise prescriptive. It is logical, therefore, that the earliest testing (e.g., Project SAILS, 2011) also focused on these skills. There are a few prominent exceptions that go beyond the mechanics of research to address the complex issues of asking a researchable question (Badke, 2011; Booth, Colomb, & Williams, 2011; George, 2008).

The second phase, the holistic writing assessment movement, was borne out of the desire to reliably judge the overall impression of a text. Based on the work of Diederich (1974), ETS took a leadership role in the development of this method that became the focus of more than a decade of writing assessment research (Elliot, 2005). In addition, using portfolios rather than essays for assessment came into fashion—a method highly compatible with the process view of writing instruction (Durst, 2006). Likewise, advocates of information literacy have called for direct, holistic assessment for similar reasons (Oakleaf, 2007). Yet there are some composition voices that call holistic scoring into question. This questioning began in the 1990s when writing assessment researchers worried that holistic scoring and the focus on reliability was crowding out understanding and validity (Huot, 1996; White, 1995). Because holistic scoring was developed for the essay, they argued, it may not be as well suited to portfolios, and they suggested that students' own reflections on their work become part of the portfolio assessment process

(White, 2005). Information literacy assessment is still largely riding Yancey's (1999) first and early second wave; limited-response tests still predominate.

2.6 Empirical Studies of Information Literacy

There were few early experimental studies of user instruction or behavior. Throughout the 1970s, a few empirical studies were conducted, although most still counted inputs. For example, Hacker (1975) counted uses of a videotaped instructional lecture. One of the earliest mixed-method studies, conducted at Earlham College, twice compared two methods of instruction, lecture versus guided exercise, using multiple measures (Kirk, 1971). The data collected included two cognitive measures: an evaluation of authentic student work, a bibliography, and scores on two objective tests of library skills, as well as an affective measure—a survey of student attitudes about the instructional program. The results showed neither teaching method was superior, but in commenting on one of the study's failings, Kirk noted that the instructors sometimes assigned questions for which materials were not available in the library. This was corrected in the second study, when the librarians researched the assignments in advance to ensure that they were researchable by the students. This early indicator points to a continuing gap in information literacy research, specifically, collaboration between faculty and librarians. Collaboration is especially prevalent among librarians working with composition courses or writing centers (Brasley, 2008; D'Angelo & Maid, 2004; Elmborg & Hook, 2005; Leckie, 1996; Norgaard, 2003; Samson, 2010), but the literature has been largely confined to theory, advice, and case studies. Several other studies of library instruction used experimental methods and quantitative analysis. Breivik (1974) and Whitmire (2002a) have been concerned with the effect of library instruction on

minority populations. Breivik's (1974) study showed a positive effect of instruction on completion rates for open admission students at the City University of New York. Seeking to assess the short- and long-term effects of instruction, researchers assessed samples of DePauw University freshmen and seniors over a three-year period. They found statistically significant effects in both short-term skill acquisition and long-term learning, while controlling for the effects of other predictors such as the number and level of courses taken as well as SAT and GPA scores. They also compared the effects of instruction to other factors such as overall intellectual ability or diligence and also showed that practice increased skill retention (Hardesty et al., 1982).

Others have quantified the effects of library instruction on student learning and transfer using multiple regression techniques. Using the work of Tinto (1975) and Astin (1984), Whitmire (2001) sought to understand if library use, by a variety of measures, increased students' critical thinking skills. She made connections between the questions asked by 'college impact' researchers and the library. She used factor analysis and hierarchical multiple regression to discover the factors that most influenced learning, finding that focused library activity had a measurable impact. Some of her studies employed qualitative techniques to tease out underlying beliefs and patterns of information seeking in undergraduates (Whitmire, 2002b, 2003, 2006).

It was not until the ACRL Standards were in place that information literacy instruction became a significant feature of academic library services, and the literature began to reflect this new interest. In 2006, Koufogiannakis published a rigorous meta-analysis of effective teaching methods in which she compared online, face-to-face, and hybrid instruction. Her extensive search of 15 databases yielded over 4,000 articles from

which she ultimately culled 17 that met her inclusion criteria. The selection criteria stipulated only that a librarian be involved in an undergraduate instruction session, and that the study utilize a measure of the cognitive outcome effect of instruction via some test, paper, or bibliography. She did not find one method more effective than another, but she did find that instruction had a measurable effect. This may seem obvious, but does provide evidence that librarians are effective teachers of information literacy to those asking for institutional support of academic libraries. Only 7% of the researchers (13 studies) had validated their instrument. The majority of these studies were limited-response tests (Bostick, 1992; Cameron et al., 2007; Cheung, 2002; Diller & Phelps, 2008; Green & Bowser, 2006; Gross & Latham, 2007; Ivanitskaya et al., 2005; McClure, Cooke, & Carlin, 2011; O'Connor et al., 2001; Ondrusek et al., 2005; Roberts, 2004; Rosenblatt, 2010; Scharf et al., 2007; Sherman, Martin, & An, 2011).

Only studies by five of these researchers, Diller, McClure, Roberts, Rosenblatt, and Scharf, used authentic student work. Roberts used a constructed response to compare the information-seeking skills of nursing students in two different programs. The four remaining authors used student work to evaluate ACRL Standard Four. McClure (2011) analyzed types of citations, frequency, and how sources were used in the final paper to study the effect of an online tutorial. She found the tutorial improved the bibliographies, but highlighted integration of sources into the text as a problem needing further study. Rosenblatt evaluated citations in 20 papers written by upper-division undergraduates and found that following library instruction 85% of the students in her sample were able to find and cite appropriate sources. She followed up with an assessment of how well students integrated sources into their texts and found 50% had difficulty doing so. The

portfolio assessment of Scharf et al. (2007) had similar findings, and evidence of "integration" resulted in the weakest scores. The studies by Scharf et al. (2007) and Diller (2008) had much in common. Scharf sampled writing portfolios that contained an entire semester's work, whereas Diller sampled portfolios that took a "best papers" approach, with students making the selection from among their assignments to provide evidence of the learning outcomes. Both used rubrics to ensure interrater reliability. Unlike Scharf, Diller asked students to write reflective essays about their learning and rated these with a rubric as well. Scharf compared outcomes on separate writing and information literacy rubrics that were rated by the instructors and librarians teaching the course, whereas Diller's readers were part of a special committee who were independent of course instruction.

2.7 Context Matters

Cronbach (1975) observed the complex nature of the interactions between characteristics of the person, the environment, and the treatment. However, the ACRL Standards refer to competencies that are divorced from any particular aptitudes, situation, or instruction. In addition, the Standards attempt to generalize and thereby reduce behavior to a set of consistent behaviors. Yet librarians have long understood the importance of context in information behavior (Buschman, 2009; Fisher, Erdelez, & Mckechnie, 2005; Gould, 1988; Gould & Handler, 1989; Gould & Pearce, 1991; Lloyd & Williamson, 2008; Warner, 2008). Thus, in practical terms, performance indicators of information literacy are commonly expressed as they are contextualized in a field of inquiry. Thus, one major problem facing researchers is consistency in the operationalization of the definition of information literacy because it must be constantly

recontextualized. This requirement also presents challenges to testing reliability and validity of measures across disciplines and institutions. Although researchers often align their measures with ACRL Standards, they may not be testing the same constructs, as illustrated by the previously mentioned comparison of the iSkills test with a locally developed direct assessment (Katz et al., 2008). Despite the face validity of each as a measure of information literacy, and although each measure correlated with a number of other measures of student performance, no relationship was found between the two measures of information literacy, which implies that the two assessment tools might be measuring different constructs. The results of Katz et al.'s study highlight an important validity issue regarding the measurement of information literacy. Cronbach's (1975) conclusion that the social scientist's task is "to articulate its generations' facts in order to gain insight into behavior" remains relevant today. In experimental work so far, information literacy researchers using authentic assignments to assess Standard Four found that context matters.

2.8 Summary

A driving force behind the current interest in information literacy research is the climate of ongoing assessment guided by the accrediting bodies in higher education. In the literature of the last 30 years, educators have frequently modeled student learning as an iterative process. In these models, learning begins with planning and goal setting, followed by implementation, data collection, evaluation, identification of needed improvements, and a return to the beginning to close the loop by reevaluating the objectives and revising the succeeding steps. This literature review began with a planning phase, by exploring the definitions and theories from librarianship and educational

psychology which are fundamental to contemporary understandings of information literacy. Instructional theories, including the cognitive and affective dimensions, were discussed. Finally, the critical tools and techniques that have been used throughout the 20th and 21st centuries for acquiring and analyzing assessment data were highlighted. The writing assessment community provided models for addressing issues of reliability and validity in evaluating process-oriented abilities that are applicable to the assessment of information literacy. Therefore, in analyzing one major theme in this literature review, we may ask the following. If information literacy assessment is the child of higher educational assessment, has it followed the prevalent assessment model? Is that model adequate for our purposes? The literature to date lacks a detailed model of information literacy assessment in higher education that can serve as a framework for the host of multivariate studies that could help illuminate pathways to improved instruction and achievement. A secondary theme followed the discussion of the quantity and validity of research that has addressed information use. Few empirical studies have addressed "use" as described in ACRL Standard Four: "[the student] uses information effectively to accomplish a specific purpose." Chapter 3 will propose such an assessment model and suggest a specific study to assess student outcomes for Standard Four and the higher-order information literacy skills.

CHAPTER 3: THEORETICAL MODEL AND RESEARCH QUESTIONS

3.1 Proposed Information Literacy Assessment Model

A "Model for Evaluating the Role of Information Literacy in Student, Program, and Institutional Assessment in Higher Education" is proposed in Figure 3-1. It describes the process of information literacy instruction and assessment and provides a framework for assessing its impact on individual students, its role in institutional effectiveness, and its impact on post-college success. Information literacy as defined by ACRL overlaps with the skills known as critical thinking. However, information literacy is skill-based, akin to math literacy or technology literacy, whereas critical thinking is a more abstract ability needed in higher-order cognitive activities in all knowledge domains. Critical thinking may be demonstrated in a variety of situations, and is a fundamental element of the thinking processes needed for information literacy (Albitz, 2007).

Measures of information literacy have included standardized and locally developed tests, either using limited-responses or constructed responses. Standardized tests are reliable, but not easily integrated into specific disciplines or content areas. Nor are they optimal for testing higher-order skills. Standardized tests have been developed primarily as tools for benchmarking across institutions. There can also be issues of practicality. Students may be more motivated to do their best and benefit from feedback if a test fits a real world instructional environment. Moreover, if an assessment can double as an assignment, it may be welcomed by faculty as an aid to instruction and course improvement, as well as making it easier for them to contribute to research on

institutional outcomes. For these reasons, this study utilized a locally developed direct assessment of authentic student work.

3.1.1 Preconditions

As has often been shown in college impact research, there are many individual and environmental preexisting conditions that affect student learning. Factors such as personal characteristics, innate abilities, and aptitudes, as well as external circumstances (e.g., SES and the educational environment in which a student was raised) can affect achievement and the student's potential for change when starting post-secondary education. The preconditions selected for study were those available through direct student report as well as through the student information system, e.g., parents' education level, high school attended, age, class standing, transfer status, and gender.

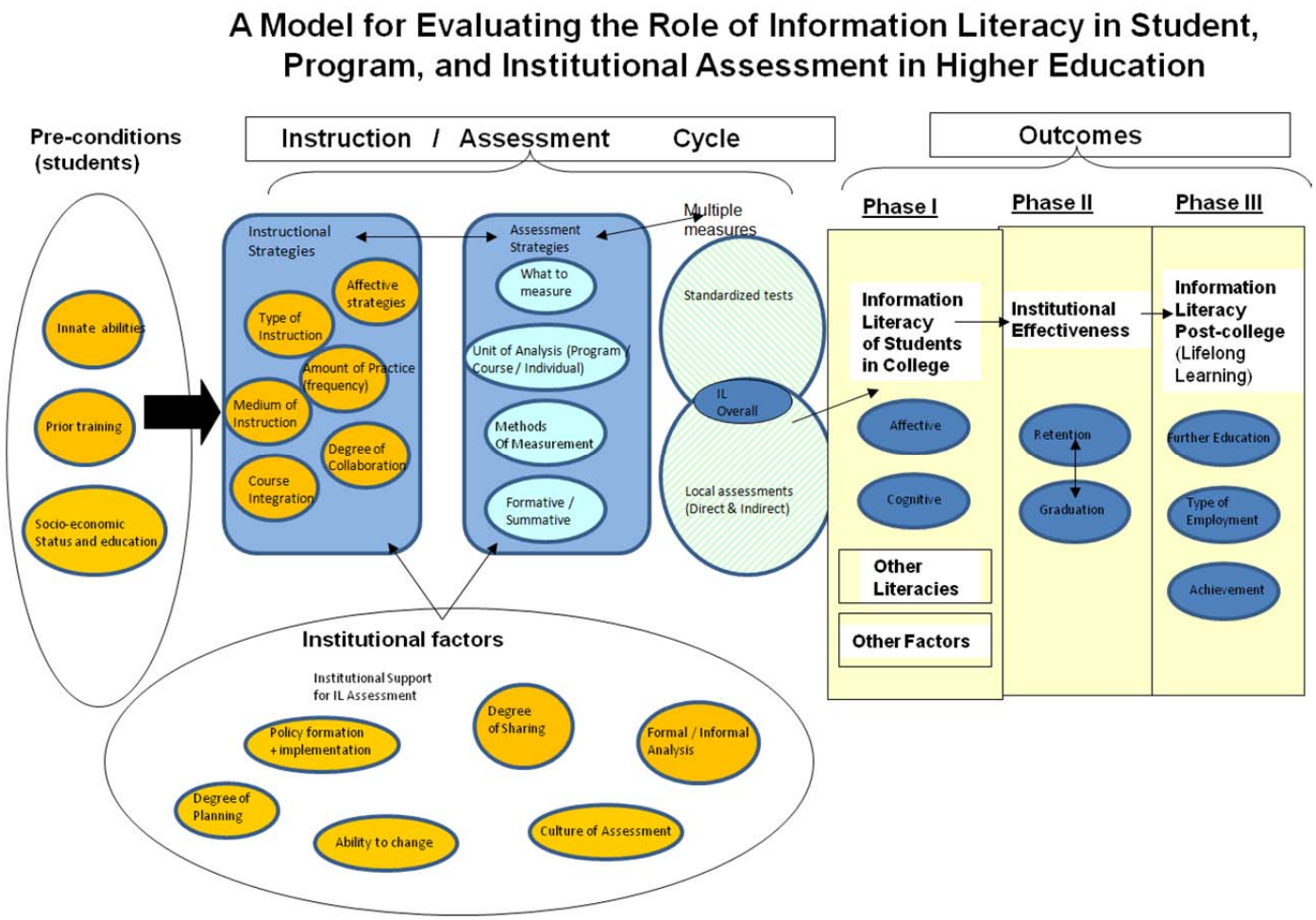


Figure 3-1. A Model for Evaluating the Role of Information Literacy in Student, Program, and Institutional Assessment in Higher Education.

3.2 The Instruction-Assessment Cycle

Assessment Strategies and Measurement

Much of the research on student learning focuses on the instruction–assessment cycle, where institutions have the best chance of accelerating student learning and achievement. There are a multitude of factors within this cycle that may affect learning. Among these are strategies such as the mode, frequency, and medium of instruction; qualities of the instructor; and the degree of collaboration between information literacy experts (i.e., teaching librarians) and other subject experts (i.e., teaching faculty). In addition, the impact of strategies that address the affective dimension of learning, such as motivation, self-regulation, or persistence, may also be measured.

The unit of analysis for assessing change or achievement may be the individual learner, the course, the program, or the institution. In the case of the individual learner, the test measures the degree of change in knowledge over time. In the case of a program or institutional assessment, the measurement may depict the ability of the program or institution to improve an aggregate measure, e.g., mean score on a standardized or local assessment of a sample of students' learning over time.

Within an academic institution, a student, instructor, program director, or administrator may use assessment techniques at different times and for a variety of purposes. These may be formative (i.e., with the purpose of improving learning) or summative (i.e., providing a benchmark of achievement at a particular point in time). The difference between formative and summative assessment is analogous to the difference between learning and achievement. In higher education, the word "learning" is frequently used interchangeably with "achievement" or "outcome." A more precise definition is that

"learning" describes a change in behavior over time, whereas "achievement" marks the accumulation of learning at a particular point in time (Shavelson, 2009, p. 11). Therefore, a learning assessment must employ data from at least two points in time and measure the change. In addition, information literacy is a multifaceted competency and may best be assessed through multiple measures using standardized nationally normed tests combined with locally devised assessment tools. Prior studies have shown that each may provide unique contributions to understanding the factors that affect learning, although they may purport to describe the same construct. These tools may include direct measurements of student performance or indirect measures such as self-reported assessments of cognitive and affective elements, both of which are thought to affect information literacy learning.

Factors Affecting the Instruction–Assessment Cycle

Institutional factors may inhibit or promote information literacy learning. The institutional culture of assessment and the level of institutional support for information literacy learning may affect the degree and nature of the institutional stakeholders' involvement. It may affect policy formation and implementation needed to sustain an information literacy program. The institutional culture and leadership may also mediate other factors that affect the instruction–assessment cycle. These might include the degree of communication across and within organizational units, or the expected degree of formal and informal analysis.

Outcomes

In Phase I of the model, individual information literacy achievement levels and learning outcomes provide good information for course and program coordinators attempting to improve teaching and learning. In addition, the outcomes serve as

indicators to individual students of their own progress and achievement. In Phase II, the aggregate measure of information literacy over time represents the average amount of learning achieved by a particular cohort of students. This can suggest policy changes that institutional administrators may employ to improve learning outcomes and institutional effectiveness (e.g., retention and graduation rates). It also indicates the effectiveness of the instruction–assessment cycle itself. This is of primary interest to accrediting agencies, whose main concern is that an institution have an ongoing process for self-evaluation and improvement in place. Institutions of higher education are increasingly being asked to benchmark their outcomes as manifested in aggregate measures of the students' success in society. What percentage of the institution's students have gone on to further education, leading to steady professional employment and good incomes? How many students demonstrate the information literacy skills of a lifelong learner 5, 10, or 20 years after graduation? Phase III of the assessment provides this. The challenge of information literacy assessment at all levels of analysis is primarily that of effect size. With so many factors inside and outside the instruction–assessment cycle, a methodology is needed to isolate the effects of interventions designed to improve information literacy learning.

3.3 Research Questions

This study employed an experimental design to study several aspects of the overall model, the effect of instruction on student learning outcomes in information literacy, while accounting for certain covariates that may influence learning. How can we determine the extent of a student's information literacy competence? How can we help students improve their abilities in this area? In an effort to shed light on these questions, this study will address the following specific questions:

1. Is a brief essay a reliable and valid instrument for assessing higher-order information literacy skills in college students?
2. Will an intervention designed by the collaborating researcher and instructor improve students' ability to use researched information effectively in their written work?
3. Which prior conditions have a significant effect on a student's information literacy performance?
4. What can we learn from how students use course materials and how they understand their own learning?

This study is unique because few prior studies have attempted to assess contextualized higher-order information literacy skills using empirical student performance data. Chapter 4 will describe the methods used to answer these questions and describe what is unique about the approach.

3.4 Evaluating Higher-Order Skills

Because this study sought to evaluate higher-order information literacy skills, a discussion of what constitutes higher-order skills is in order. Bloom's *Taxonomy of Educational Objectives* (Bloom et al., 1956) is used as a framework for understanding the complexity of cognitive abilities and their relationships to each other. Bloom first developed the taxonomy as a classification system for creating consistency in communication of educational goals and objectives (and thus standards) across institutions, disciplines, and educational levels. This common language was seen as a necessary basis for the development of curriculum, instruction, and evaluation. The taxonomy addresses cognitive, affective, and psychomotor dimensions of learning. The

cognitive process is broken down into a hierarchy of six categories in which the skills become increasingly complex. compares Bloom's *Taxonomy* with Krathwohl's revised model showing the lower order (remember, understand, and apply) and the higher-order skills (analyze, evaluate, and create) (Callister, 2010). Krathwohl's (2002) categories and subcategories were the basis for development of the instrument and rubric.

Although widely used in the field of education, there are only a few mentions of Bloom's Taxonomy in the library literature on information literacy. The ACRL Standards for Information Literacy state that lower- and higher-order skills based on Bloom's taxonomy are evident throughout the Standards, and two examples are given, but no further elaboration or systematic mapping is made. Librarians were advised by Radcliff et al. (2007) that the Taxonomy can provide guidance, and Callister (2010) proposed a mapping of legal research skills to the Taxonomy, but did not use the ACRL framework. A model of information literacy developed at the University of Worcester, UK, acknowledges the broad categories of the well-established frameworks of information literacy mentioned earlier, SCONUL from the UK, and ACRL from the United States, as well as the Australian and New Zealand Information Literacy (ANZIL) framework, without specifically referencing any one in detail (Keene, Colvin, & Sissons, 2010). Colvin and Keene (2004) had identified specific skills as part of the research process in a study that employed a student questionnaire for data collection. However, their work did not include an analysis of student work products. Therefore, a mapping of ACRL Standards to Krathwohl's revision of Bloom's Taxonomy is proposed in Table 3-2. It is evident that most of the information literacy competencies and outcomes are among the higher-order skills.

Table 3-1. Bloom's Taxonomy of Educational Objectives and Krathwohl's Revision

Bloom's Taxonomy (left), Krathwohl's Revised Taxonomy (right)

Structure of the Original Taxonomy	Structure of the Cognitive Process Dimension of the Revised Taxonomy
1.0 Knowledge	*1.0 Remember*—Retrieving relevant knowledge from long term memory.
1.10 Knowledge of specifics	*1.1 Recognizing*
1.11 Knowledge of terminology	*1.2 Recalling*
1.12 Knowledge of specific facts	*2.0 Understand*—Determining the meaning of instructional messages, including oral, written, and graphic communication.
1.20 Knowledge of ways and means of dealing with specifics	*2.1 Interpreting*
1.21 Knowledge of conventions	*2.2 Exemplifying*
1.22 Knowledge of classifications and categories	*2.3 Classifying*
1.24 Knowledge of criteria	*2.4 Summarizing*
1.25 Knowledge of methodology	*2.5 Inferring*
1.30 Knowledge of universals and abstractions in a field	*2.6 Comparing*
1.31 Knowledge of principles and generalizations	*2.7 Explaining*
1.32 Knowledge of theories and structures	*3.0 Apply*—Carrying out or using a procedure in a given situation
2.0 Comprehension	*3.1 Executing*
2.1 Translation	*3.2 Implementing*
2.2 Interpretation	*4.0 Analyze*—Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose.
2.3 Extrapolation	*4.1 Differentiating*
3.0 Application	*4.2 Organizing*
4.0 Analysis	*4.3 Attributing*
4.1 Analysis of elements	*5.0 Evaluate*—Making judgments based on criteria and standards.
4.2 Analysis of relationships	*5.1 Checking*
4.3 Analysis of organizational principles	*5.2 Critiquing*
5.0 Synthesis	*6.0 Create*—Putting elements together to form a novel, coherent whole or make an original product.
5.1 Production of unique communication	*6.1 Generating*
5.2 Production of a plan, or proposed set of Operations	*6.2 Planning*
5.3 Derivation of a set of abstract relations	*6.3 Producing*
6.0 Evaluation	
6.1 Evaluation in terms of internal evidence	
6.2 Judgments in terms of external criteria	

Table 3-2. ACRL Information Literacy Standards Mapped to Krathwohl's Revision of Bloom's Taxonomy

ACRL Information Literacy Standards Mapped to Krathwohl's Revision of Bloom's Taxonomy

The Knowledge Dimension	The Cognitive Process Dimension					
	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual Knowledge		5.1. Understands many of the ethical, legal and socio-economic issues surrounding information and information technology (Ideas)	1.2. Identifies a variety of types of formats of potential sources of information (Evidence of Research)	3.4. Compares new knowledge with prior knowledge to determine the value added, contradictions, or other unique characteristics of the information (Ideas)		
Conceptual Knowledge				1.1. Defines the need for information (Evidence of Research)	1.4. Re-evaluates the nature and extent of the information need (Improvement in post-test)	1.1. Articulates need for information (Writing)
				1.3. Considers the costs and benefits of acquiring the needed information (Evidence)	4.2. Revises the development process for the product or performance (Improvement in post-test)	3.3. Synthesizes main ideas to construct new concepts (Writing)
				3.1. Summarizes the main ideas to be extracted from the information gathered (Ideas and Writing)		4.1. Applies new and prior information to the planning and creation of a particular product or performance (Ideas and Writing)
						4.3. Communicates the product or performance effectively to others (Writing)
Procedural Knowledge			2.3. Retrieves information online or in person using a variety of methods (Evidence of Research and Citation)	3.2. Articulates and applies initial criteria for evaluating both the information and its sources (Evidence of Research)	2.1. Selects the most appropriate investigative methods or information retrieval systems for accessing the needed information (Evidence)	2.2 Constructs and implements effectively-designed search strategies (Evidence)
			2.5. Extracts, records, and manages the information and its sources (Citation)		2.4. Refines the search strategy if necessary (Evidence of Research)	
			5.2. Follows laws, regulations, institutional policies, and etiquette related to the access and use of information (Citation)			
			5.3. Acknowledges the use of information sources in communicating the product or performance (Citation)			
Meta-Cognitive Knowledge					3.4. Compares new knowledge with prior knowledge to determine the value added, contradictions, or other unique characteristics of the information. (Integration)	3.6. Validates understanding and interpretation of the information through discourse with other individuals, subject-area experts, and/or practitioners (peer-review) (Integration)
					3.5. Determines whether the new knowledge has an impact on the individual's value system and takes steps to reconcile differences (Integration)	
					3.7. Determines whether the initial query should be revised (Improvement in post-test)	

KEY: Numbers refer to ACRL standard and performance indicator. ACRL Standard 1=pink; ACRL Standard 2=green; ACRL Standard 3=yellow; ACRL Standard 4=blue; ACRL Standard 5=grey

This research was also informed by John Biggs's theory of constructive alignment. Whereas Bloom's work took the perspective of the teacher and focused on methods of instruction that maximized learning, Biggs focused on what and how students learn (Biggs, 1999/2011). Constructive alignment dictates consistency in learning goals and assessments. It rests on criterion-referenced assessment and employs learning activities that embody the assessment criteria. Biggs shows that practicing constructive alignment requires teachers to communicate the learning goals and the criteria for assessment to the learners. This combination of consistency and communication enables students to see the connection between the goals, the content, the activities, the assignments, and the evaluation of their learning through assessment of their work product, thereby aiding their learning.

CHAPTER 4: METHODS

4.1 Introduction and Overview

The four research questions were addressed by data collected during a study conducted in the spring and summer 2011 at the New Jersey Institute of Technology (NJIT) among students enrolled in an upper-division course on technical communication. This chapter will discuss the participants, instrument, intervention, procedures, and the choice of study method for each research question, including the statistical methods chosen for data analysis. The researcher is a librarian at NJIT who collaborated with a course instructor on the curriculum development, the sequencing of readings and assignments, and implementation. The students in what is called here "the treatment group," were provided with an intervention consisting of multi-media materials and assignments provided online through the course management system designed to teach, practice, and assess component skills associated with information literacy. In order to evaluate the effect of integrated information literacy instruction on student learning outcomes among these students, the following steps were taken. A curriculum was designed, implemented, and assessed. To design the curriculum, educational objectives were first established and an instructional intervention and an assessment protocol were developed. This study also benchmarked student performance in sections of the course taught by other instructors (who did not participate in the information literacy treatment) (i.e., the control group). The instructors of those sections constructed their own instructional plans and assignments without collaborating with the librarian.

A new instrument for assessment, the diagnostic essay, was developed and tested. It is a locally developed direct assessment in the form of a brief researched essay that was

validated and used to measure several components of information literacy (i.e., the dependent variables). It was employed as a pretest (i.e., formative assessment) and posttest (i.e., summative assessment) of learning. Institutional Review Board approvals from Rutgers University and NJIT were obtained. Data collection began with the administration of a survey and consent form to participants. Data was collected online via a survey tool and the online course management system in use at NJIT, Moodle. The pretest was administered to the students in the treatment group. Following the instructional intervention, students were given the same essay assignment as a posttest of learning. Those in the control group did not take the pretest, nor did they receive the intervention, but they did take the posttest at the end of the semester. The collaborating instructor was responsible for day-to-day course management and grading. The researcher was available for consultation with students and handled data collection, management, and analysis.

The variables for the study were selected from among the major components of the proposed model: preconditions, institutional factors, instructional strategies, assessment strategies, and instruments for measurement. Preconditions selected for study are those variables available through direct student report as well as the student information system: Parents' education level, high school attended, age, class-standing, transfer status, and gender. Institutional factors are not a subject of direct study in this research project. However, the work took place at an institution where the information literacy abilities of graduates have been a part of the formal goals of the university since 2009 (New Jersey Institute of Technology, 2009). Also, since 2005, instruction and assessment of these abilities in undergraduates have been ongoing activities of both the

library and the department in which the study was conducted. Barriers to the methods used in this research study might present insurmountable obstacles at institutions lacking this school's commitment to and experience with information literacy instruction and assessment. In addition, because this study measures individual student learning gains within one course in one university, the institutional factors are presumably fairly constant. Furthermore, because an existing course for this study was selected, instructional strategies and instruments for measurement were necessarily constructed using prior research at this institution.

The theories on which the study protocols were developed, and the details of study implementation are discussed in the following sections. Figure 4-1 highlights the project's zones of study.

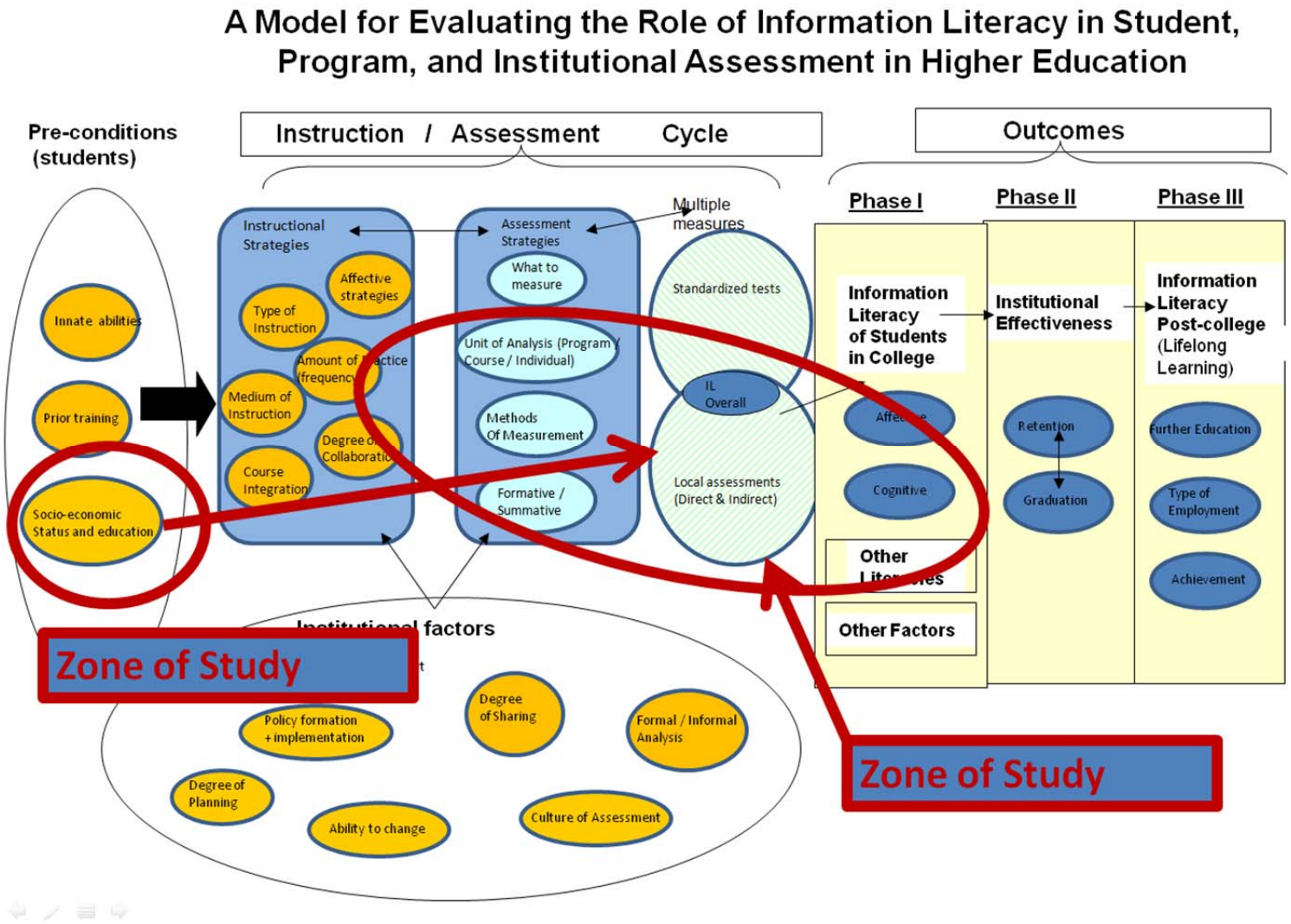


Figure 4-1. Zone of Study in a Model for Evaluating the Role of Information Literacy.

4.2 Participants and Context

This section provides a description of the students, faculty, and the environment in which the experiment was conducted during the spring 2011 and the summer 2011 semesters. NJIT is a New Jersey public university founded originally as the Newark Technical School in 1881. It has 47 bachelor's programs, 59 master's programs, 19 doctoral programs, 490 full and part-time faculty, and an enrollment of approximately 6,500 undergraduates and 3,000 graduate students. NJIT is one the most ethnically diverse colleges, but the undergraduate population gender is skewed with approximately 75% of the undergraduate student body being male. The study participants were 20% female and 80% male. None were freshmen, 4% were sophomores, 43% were juniors and 53% were seniors. The average SAT scores for undergraduates for fall 2010 were 603 Math and 537 Verbal. Table 4-1 shows the breakdown of GPA by gender for fall 2010.

Table 4-1. Average GPA for Study Participants Compared to NJIT Undergraduates Fall 2010

	Study Participants		**All Undergraduates**		**Seniors Only**	
Gender	#	Avg GPA Jan 2011	#	Avg GPA F2010	#	Avg GPA F2010
Female	31	2.94	1356	2.91	320	3.01
Male	131	3.02	4747	2.76	1235	2.88
All	162	2.98	6103	2.79	1555	2.91

Source: NJIT Web site Enrollment data

The institution is highly focused on science and technology, thus a majority of students major in science, math, engineering, and computer science. In 2011, for example, 69% of the undergraduate degrees were granted in engineering- and computer-science–related majors (New Jersey Institute of Technology, 2012). According to a survey of entering

students conducted in fall 2010, 57% of the undergraduates had at least one parent who graduated from college (Deess, 2010). Over 30% of NJIT undergraduates are transfer students. For example, the percentage of new incoming undergraduates who were transfer students fluctuated between 2009 and 2011 from 38% to 55% to 48% respectively (National Center for Education Statistics, 2012).

The demographic profile of the students in this study is relatively close to the profile of degree-seeking undergraduates compared to data compiled in the two academic years prior to this study (when most of these students presumably began their studies at NJIT). The percentage of African Americans and Latino students enrolled in our study was slightly lower and the percentage of Caucasians slightly higher than the overall profile of NJIT students. See Table 4-2 for percentages.

Table 4-2. Ethnicity Compared to NJIT Data

		NJIT Degree Seeking Undergraduates	
Ethnicity	**Study Participants**	**Fall 2009**	**Fall 2008**
Asian	22%	21%	21%
African-American	4%	10%	10%
Hispanic	16%	20%	19%
International	2%	4%	5%
Native	1%	1%	1%
Caucasian	48%	36%	35%
Unknown	7%	8%	9%

The course selected for study is an English course on technical communication, which is offered regularly by the Department of Humanities. It is an upper-division course requirement for all STEM (science, technology, engineering, mathematics) majors at NJIT, where such majors represent approximately 85% of the undergraduates (New Jersey Institute of Technology, 2012). Thus, the students enrolled in this course represent

a reasonable and fair-sized sample of typical undergraduates at the institution. Writing instructors who teach this course and their liaison librarian were already collaborating. It was hypothesized that a nontraditional writing assignment might better motivate students and integrate course objectives in both writing and information literacy. The assignment would be designed to motivate students to expand their research skills while addressing all five of the ACRL Standards. The curriculum development evolved over several semesters with the librarian and three different instructors. This course was selected for this pilot because the course goals encompass writing, speaking, thinking, and research objectives. Previous portfolio assessment in the technical communication course had pointed to weaknesses in the students' ability to cite their sources, but the researchers did not consistently measure this criterion (Johnson, 2006). Typical assignments in this course may involve writing instructions, documenting a personal interview, writing a resume, or making a proposal using PowerPoint. Students learn to analyze complex communication situations and design appropriate responses through tasks that involve problem solving, rhetorical theory, document design, oral presentations, writing teams, audience awareness, ethical considerations, and gender equity issues. The course format is not a subject of investigation, because no particular course format (online, face-to-face, or hybrid) has been shown to definitively affect learning (Means, Toyama, Murphy, Bakia, & Jones, 2009).

4.3 Learning Objectives and Assessment Criteria

As described in the review of the assessment literature, the first step in the instruction-assessment cycle is the development of learning objectives. The course objectives include the following information literacy competencies (presumably taught in

all sections): 1) find and evaluate sources for quality and relevance; 2) cite sources; 3) use sources for a specific purpose. The librarian had previously established a collegial and collaborative relationship with all the instructors and they were aware of the information literacy objectives.

The curriculum and instructional plan flow from the learning objectives. However, as Tyler (1949) and Bloom (1968) understood, curriculum and instruction are considered in relation to the questions that must be answered about the effectiveness of the teaching program. In other words, the methods of assessment must be considered in order to plan the curriculum and instruction. Thus, it is common to formulate assessment criteria, and devise the curriculum, instruction, and an instrument (test or assignment) for use with the criteria. The overarching goal of the course used in this study is for students to be able to effectively craft a communication package that accomplishes a specific goal in a given situation for a specified target. To do so, they are expected to seek and evaluate information and use it to inform and support their writing. The assessment criteria were developed followed the ACRL Standards and are described in detail below. They are also shown in parentheses as they relate to Bloom's *Taxonomy* in Table 3-1.

Evidence of Research and Quality of Sources (Evidence). This criterion was designed to capture the student's ability to know when information was needed (ACRL Standard One), identify and access sources (ACRL Standard Two), and select sources appropriate to the task (ACRL Standard Three3). A rating of "very poor" is given if the student fails to go beyond the sources given in the assignment prompt and indicates no evidence that any additional sources were sought. On the opposite end of the scale, an essay would receive a rating of "accomplished" if the work demonstrates that a student

understood that additional research was needed by the use of several good quality sources appropriate to the task at hand.

Mechanics of Citation (Citation). This trait is intended to isolate a student's ability to use academic conventions in citing the sources used (ACRL Standard Five). In prior studies at NJIT (published and unpublished) this trait had proven to be a good proxy for overall information literacy performance (Scharf, et al., 2007). Scores range from "very poor," where citations are missing or only a URL or an author and title are present, to "accomplished" where all sources were fully and correctly cited in APA format both in the text and the bibliography.

Ideas/Integration of Sources into Content (Ideas). This trait is designed to capture the use of information for a purpose (ACRL Standard Four). Readers are asked to read for meaning to see if the student offered evidence from outside sources to support their argument, reasons, or explanations. The essays receiving higher scores use sufficient evidence appropriately and effectively to convince.

Writing. The researchers found no published studies besides those at NJIT where the quality of the writing was assessed in relation to information literacy (Elliot, Briller, & Joshi, 2005; Elliot et al., 2007; Elliot, Kilduff, & Lynch; Johnson & Elliot, 2004; Katz et al., 2008; Scharf et al., 2007). In prior work, it was discovered that writing ability, especially in persuasive writing, was an important factor in the quality of the thinking, and highly related to the students' ability to use the information they gathered (Scharf et al., 2007). This trait was used to collapse both the form and content attributes usually associated with writing into a single score. Scorers were instructed to rate essays lower

based on the seriousness of the mechanical errors, lapses of logic, disorganization, poor word choice, or lack of clarity.

Holistic and Sophistication (Holistic). Although the other traits are analytical, this holistic trait allowed scorers to make an overall judgment on the ability of the student to use information effectively, while taking the level of sophistication into account. If, for example, a student communicated the central idea clearly, but exhibited a low level of complexity in thinking, readers were advised to rate the essay as weaker. They were likewise advised to rate the essay as weaker if a student responded clearly to the assignment, but employed a minimal use of sources to support the argument(s).

4.3.1 The Instrument

An instrument was developed that consisted of an essay prompt designed to test higher-order skills and elicit student self-reflection about her or his process. Several elements underlie the choice of an essay as a diagnostic instrument. A diagnostic essay as a form of test that requires a constructed response was chosen over a limited-response test for its ability to test higher-order skills and be more easily integrated into specific disciplinary or content areas. Limited-response and standardized tests have typically been employed for their reliability, and as benchmarking tools for benchmarking institutions. In addition, the use of standardized tests often raises issues of practicality for the instructor, and challenges test-taker motivation. On the other hand, a brief essay test can fit more easily into a real-world instructional environment, and students may be more motivated to do their best and benefit from formative feedback. Instead of anonymously checking a box, students are speaking in their own voices, and thus may put more care and effort into this representation of themselves. Guessing is not part of the equation. In

teaching and assessing complex integrated skills that involve critical thinking, writing, and research, limited-response tests can also be limited in range and depth. Furthermore, if an assessment can double as an assignment, it may be welcomed by faculty as an aid to instruction and course improvement, as well as make it easier for them to contribute to institutional outcomes research.

Bloom suggested that the presentation of a novel situation using a constructed-response format is appropriate for testing higher-order skills; the new instrument conforms to Bloom's suggested examples for the testing of such skills. What was presented to the students as a topic for the essay concerned issues of privacy and the use of social media in hiring and maintenance of employment, a topic it was hoped would be engaging to most college students. The assignment called for a persuasive essay. The most persuasive answers would employ arguments supported by information in addition to the sources mentioned in the prompt. The second part of the essay called for a self-reflective statement on the process of seeking information and responding to the prompt. The assignment served as the instrument for assessment. The situation was presented as follows:

> As social networking becomes more common, employers have begun to review the pages of current and prospective employees though it may not be a valid measure of work habits or attitude. In the article "No Place to Play: Current Employee Privacy Rights in Social Networking Sites," Genova (2009) takes the view that employers have a right to monitor employees' information on Facebook. At the same time, many of those seeking employment do not consider this when posting details of their

> own lives on these types of pages. "Examining Students' Intended Image on Facebook: 'What Were They Thinking?!'" an article recently published in the *Journal of Education for Business* indicates that people may be disregarding the possible negative consequences of what they post on Facebook. Beyond this, studies have shown that many new employees are spending time on Facebook during the workday without any work-related reason for this access, negatively impacting productivity. Given these issues, do you feel that your activity and information on social networking sites like Facebook should have an impact on your ability to secure a job or maintain your employment?

The students were asked to submit a formal research essay responding to the question listed above in at least 300 words, as well as a reflective statement. They were told that the reflective statement should explain the thinking and process used to write the essay by explicitly describing how they engaged with the topic, how they found material to support their arguments, and how they determined their sources to be reliable, trustworthy, and of high enough quality for the essay. They were also told that the reflective statement should be as long as necessary. (See Appendix B).

This type of instrument was also chosen because the similarities between writing and research processes suggested that similar assessment tools might also be appropriate. Essays as part of high-stakes summative pre-college testing and writing placement have long been in use and the nature of both reliability and validity issues have been thoroughly examined (Messick, 1994). At the same time, the "authentic assessment" movement, as exemplified by the work of Grant Wiggins (1989, 1990), was founded on

the desire of the composition community to capture the complexity and open-ended nature of writing in the local context. The composition community has discussed a variety of criteria that constitute "authentic" assessment and these were classified by Frey, Schmitt, and Allen (2012). According to their definition, "authentic" is not a claim to validity, but a method of assessment. Frey grouped the criteria into three dimensions—context, role of the student, and scoring method.

Context of the assessment

- realistic activity or context
- the task is performance based
- the task is cognitively complex

Role of the student

- a defense of the answer or product is required
- the assessment is formative
- students collaborate with each other or with the teacher

Scoring

- the scoring criteria are known or student developed
- multiple indicators or portfolios are used for scoring
- the performance expectation is mastery

According to these criteria, an assessment such as the one proposed here would be "authentic." The similarities between writing and research also led to consideration of a way to tease out the relationship between the two competencies. Thus, the researcher decided to join the composition and information literacy variables in one instrument. (See Appendix B for the essay question and full scoring rubric using the assessment criteria.)

4.4 The Scoring Rubric

A rubric was developed by the researcher based on an earlier assessment model in which writing and information literacy each had separate scoring rubrics. The assessment model followed by Scharf, et al. in 2007 used two sets of scores, one set for writing and one set for information literacy. This study used a revised model depicted in Figure 4-2 to simplify the scoring process while integrating writing and information literacy into a single rubric, because the earlier study suggested that writing and research are highly integrated activities.

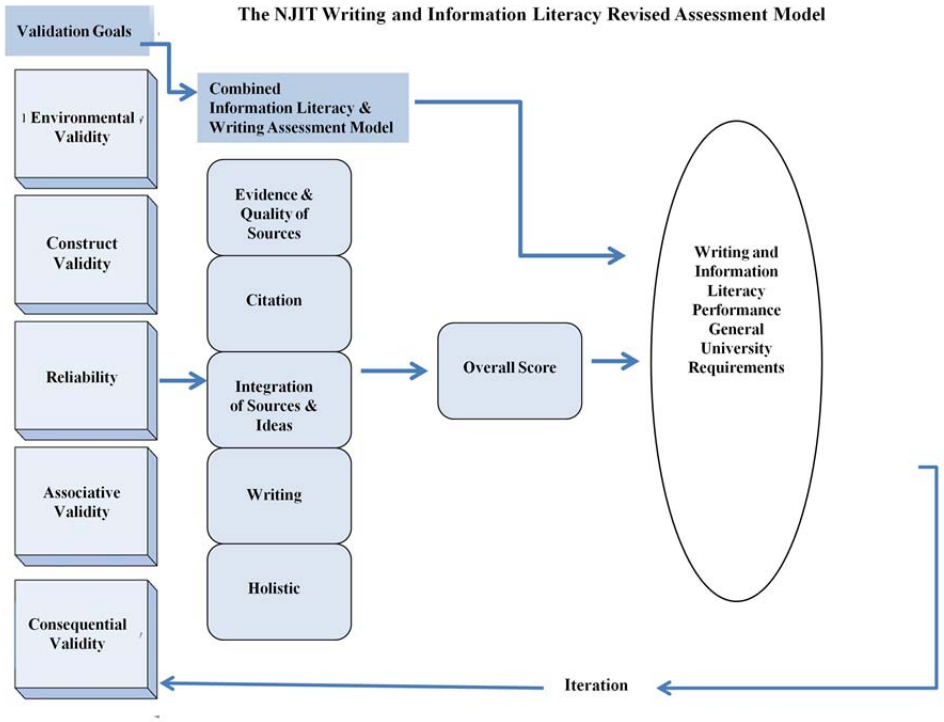


Figure 4-2 NJIT Writing and Information Literacy Revised Assessment Model

A ten-point Likert scale was used to score the essays. The rubric included both analytic criteria and a holistic score for information literacy, that is, the overall

impression it creates of the student's ability to find and communicate information effectively. The holistic scoring approach was developed by the Educational Testing Service (ETS) (Diederich, 1974) and applied widely in writing (Durst, 2006). In addition to the holistic score, individual assessments of component skills were scored using specific performance criteria as indicated on the rubric. By using this method, the researcher sought to uncover the relationships among the underlying component skills that resulted in improvement in overall information literacy. Both types of criteria were used to triangulate the information literacy scores and increase the ways in which the validity of the instrument may be judged. (See Appendix B for the text of the essay prompt and the scoring rubric.) Detailed procedures for training readers and scoring essays is discussed in section 4.7.

4.5 The Intervention

4.5.1 Sequencing of Instruction and Assignments

This section describes the sequencing of instructional content and assignments related to the information literacy intervention. Information literacy skills were threaded throughout the semester for the treatment group, culminating in the Wikipedia Improvement by Supported Expert Revision (WISER) Project. The project tasks required students to identify articles in Wikipedia that needed revision and then improve them by editing definitions, adding content, and adding reliable and appropriate sources as references. Students prepared a written proposal in which they made a case for why the changes were needed, why the student was qualified to make them, and why the sources they cited provided reliable evidence and adequate support. After revising the article, the students reflected on their work by persuading the professor in a letter that they had

accomplished the proposed work. The pretest was part of Task 1; Tasks 7, 8, 10, 11, 13, and 14 required competencies that make up the information literacy construct. Task 13 was the posttest. The treatment group followed a scaffolded course design as they acquired the competencies required in the final project. Table 4-3 indicates the sequence of tasks during the semester. In contrast, the control group was given the privacy essay as a stand-alone information literacy assignment, which was worth 5% of the course grade. The study did not control instructional models in the control group, but allowed them to vary according to the instructor, as they would in real life.

Table 4-3. List of Tasks for the Treatment Group. Highlighted rows were explicitly part of the information literacy intervention.

Due	Assignment	Value
Week 1	Task 1: Basic Tools for Class	2%
Week 2	Task 2: Introduce Yourself to the Class	1%
Week 3	Task 3: What is Communication?	1%
Week 4	Task 4: Your View of Technical Communication	1%
Week 5	Task 5: Initial Technical Communication Sample—Your Resume	2%
Week 6	Task 6: The Importance of Words and Critical Thinking	1%
Week 7	Task 7: Researching, Decision Making, and Evaluating	1%
Week 8	Task 8: Finding and Evaluating Sources	10%
Week 9	Task 9: Types of Communication Packages	10%
-	Spring Recess; No Classes Scheduled; University Open	-
Week 10	Task 10: WISER Project Proposal	15%
Week 11	Task 11: WISER Project Peer Review	10%
Week 12	Task 12: Resume Package Elements Explanation Memo	15%
Week 13	Task 13: Persuasive Research Essay	20%
Week 14	Task 14: Final Showcase Submissions	10%
Week 15	Task 15: Final Thoughts and Grade Argument	1%

The first six weeks of the course introduced the student to basic ideas about communication and critical thinking, which would be needed to accomplish the major assignments. The students were given low-stakes assignments designed to make them think in practical terms about the concepts presented in the introductory readings and

lectures by posting to an online forum. Students were required to discuss questions raised by the instructor's prompts, as well as to the online forum posts of other classmates. The topics covered in this manner during the first six weeks were:

- Definitions and how to recognize and create valid and good ones;
- Communication as a problem-solving tool; constructing meaning;
- The meaning and relationship of goals, objectives, outcomes, and success;
- Communication as a package for a particular goal, target, and situation;
- Critical thinking, argumentation and persuasion; judging quality;
- Word use, mood, style, and tone.

The pretest was taken in Week 1. The first information literacy assignment, Task 7, required students to read the librarian's method of searching for a topic, then reviewing three Websites as if they had been found as support for a paper on global warming. Students were asked to discuss the quality and reliability of the content on each page and the quality of each source. They were given guidelines in "Questions to Ask" when evaluating sources. After the assignments were submitted, the librarian, who has been positioned as the research and source expert to the students, revealed the expert's answer.

The next unit introduced material that would be needed to execute Task 8 and the WISER project (i.e., Tasks 10 and 11) effectively. It covered:

- Decision making;
- Data, information, knowledge;
- Dealing with uncertainty;
- Demonstration of Web searching for reliable results;
- Links to tips on advanced Google searching-from a Princeton librarian;

- Finding the full text of an article;
- Finding e-books;
- Understanding citations.

In Task 8, students were asked to write a written evaluation of two of the sources used to support statements made in the Wikipedia article on coaxial cable. The researcher/librarian had selected the Wikipedia article and four sources to give students the option of choosing which sources they would investigate. The sources were of varying difficulty and quality. The criteria for evaluation were provided as guidance. After students submitted their work, the librarian's expert answer was revealed to them.

The WISER Project (Tasks 10, 11) required students to identify articles needing revision, and then improve them by editing definitions, adding content, and adding reliable and appropriate sources as references. Each student chose a Wikipedia article and posted his or her choice in an online forum. In the post, they were to provide a properly formatted (APA style) citation for the article they chose, define the topic (as given in the article) and assess its quality. Then they discussed why they selected the topic, how it related to their major, and why it needed expert support and revisions. Finally, they were to include a citation for a source they might use to support the changes. The instructor and librarian monitored the online forum and intervened with feedback on topics and sources for all students to see.

The librarian provided instructional resources:

- Introduction to WISER
- WISER—example of a good page
- Librarian's introduction to Wikipedia

- How the librarian chose a topic
- Parts of a citation
- American Psychological Association Publication Manual

Next, students prepared a formal written response to the request for a proposal (RFP) to improve Wikipedia in which they made a case for why the changes were needed, why they were qualified to make them, and why the sources they proposed to use were of good quality and would provide adequate support. Students were given sample proposals as models and the information literacy rubric (see Appendix B) for guidance. Task 11 required a peer review of each student's proposal. Students submitted the final version of their response to the RFP, in which they requested permission to proceed with the work. Students then revised the Wikipedia article online. Following the WISER Project, the students took the posttest, which was the aforementioned essay on privacy, social media, and the workplace. Students then created an e-portfolio of their work, the "final showcase," which was intended to demonstrate that they had met the goals set out in the mission for the class. They were also asked to reflect on their work by persuading the instructor in a formal letter that they had actually accomplished the work they had proposed. (See the librarian's instructional materials in Appendix D.)

4.5.2 Underlying Theories of Education

The benefits of using Wikipedia as a teaching tool have been noted by some educators (Cummings, 2009; Konieczny, 2007, 2012; Jullien, 2012). Many find that students are more engaged when writing for an authentic global audience, working collaboratively, interacting with readers and writers in real time, acquiring knowledge, and gaining a deeper understanding of media literacy (Wikimedia Foundation, 2012).

They also acquire an appreciation of good documentation and a deeper understanding of the importance of context. These benefits may also be seen as evidence of the successful application of some underlying educational theories. This section will elaborate on the educational theories underlying the choice of Wikipedia and the other choices made when designing the intervention and techniques used in this study. Tyler's (1949) curriculum design, Bruner's (1960) notion of instructional scaffolding, Knowles's (2005) andragogy, student engagement and reflection, and the pedagogies of active learning, guided the instruction plan. In addition, the close collaboration between the faculty member and the researcher/librarian was an important component.

4.5.3 Curriculum Design

Tyler's (1949) work on the interaction of curriculum design, instruction, student behavior, and assessment has been particularly influential in the design of this intervention. The four chapter headings of his book summarize the questions that drove his theory. Chapter 1: "What educational purposes should the school seek to attain?" This requires a clear definition of goals. In this study, the learning goals were clearly defined and communicated to students in the treatment group. At the same time, the process of establishing goals and finding ways to attain them was also part of the course content. Chapter 2: "How can learning experiences be selected which are likely to be useful in attaining these objectives?" This question promotes thinking about what types of tools, in the form of instructional materials and assignments, would be useful in achieving the goals. Tyler uses the term "learning experience" to describe the result of the interaction between the learner and the tools and techniques that create the learning environment. Tyler believed that the learner learns through his or her own active behavior within the

learning environment, rather than by what the teacher does. Naturally, the teacher has the most important role in creating that environment. Tyler's belief was explored in more detail by educators several decades later and has become widely known in higher education as "active learning" (Bonwell & Eison, 1991). Pascarella and Terenzini's (2005) major review of the college persistence literature also suggests that positive effects occur when active learning techniques are employed. In this course, the intention of the researcher and instructor was to influence the students' learning experience by creating tasks and situations that would stimulate engagement and behaviors that would enable the students to achieve the objectives, and by positioning the instructor and librarian as coach and consulting expert, respectively, rather than authorities. Chapter 3: "How can learning experiences be organized for effective instruction?" This places the emphasis on continuity, that is, the sequencing and integration of the course components. In following Tyler's thinking, the students in the treatment group were given recurring opportunities to practice component information literacy skills. Chapter 4: "How can the effectiveness of learning experiences be evaluated?" Tyler believed in the importance and utility of continuous assessment and revision. He was neither a librarian nor a composition teacher, but since writing and research are both iterative processes, they fit well with his theory. This is an important assumption underlying the entire study. In addition, the concept of evaluation, self-criticism, and revision for quality improvement was an ongoing theme in the course content.

4.5.4 Scaffolding.

Instructional scaffolding is another fundamental concept that influenced the design of the intervention. The technique has been traced to Bruner (Wood, Bruner, &

Ross, 1976) building on Vygotsky's work. In scaffolded instruction, an expert provides initial supports for learning that enable the learner to build component skills within his or her proximal zone of development. This allows the learner to complete tasks that are initially out of reach; as the learner gains mastery, the scaffolding is gradually removed. It has been suggested that scaffolding is particularly useful for learning higher-level skills (Rosenshine & Meister, 1992). Widely known among K–12 educators, scaffolding is associated with Carol Kuhlthau's concept of "guided inquiry" (Kuhlthau et al., 2007). However, it has rarely been mentioned as a technique in the literature on information literacy. Two studies discussed use of the technique in detail, but performed no assessment of the results (Niedbala & Fogleman, 2010; Walton & Archer, 2004).

4.5.5 Andragogy

The approach taken to teaching and learning is based on assumptions about lifelong learners articulated by John Knowles (Knowles, Holton, & Swanson, 2005) in his work on the adult learner. In contrast to theories of pedagogy, Knowles's idea of andragogy begins with the following assumptions:

1. Learners need to know why, what, and how they are learning.
2. They have an autonomous, self-directed self-concept. They feel responsible for their own decisions and resist being directed.
3. They have a variety of experiences that must be valued in the learning process, but which also may present challenges due to preexisting ideas and mental models.
4. They must be ready to learn. The timing of the subject matter and the needs of the learners coincide.

5. Their orientation to learning is problem centered and contextual.
6. They are motivated to learn when they perceive it has intrinsic value or will be useful.

These assumptions are appropriate to students at NJIT, who are adults, and it influenced my approach when facilitating their learning. It required preparing learners for learning by providing information about the content and conditions, diagnosing needs, and setting learning objectives that made sense to students. Such assumptions also promoted the development of problem-based learning tasks and assignments designed to be seen as interesting, relevant, and useful. Collaboration, peer review, and reflection were also components of the instructional strategy based on these assumptions. Sharing the learning objectives and rubric for evaluation with students supports Knowles's first principle. This enables adult students to gain a clear idea of what they are supposed to be learning and places the responsibility for self-evaluation on them. This is also in line with theories of self-regulation and social feedback underlying instructional design, opportunities for self-assessment, and peer feedback, which were part of the teaching strategy.

4.5.6 Engagement

Although Knowles viewed engagement as coming from the learner, it is has been well documented that one of the most important features of superior teaching (in the student's view) is the teacher's ability to stimulate interest (Feldman, 1976). Researched writing requires engagement. The words we use make it clear that the person is "engaged in research." Yet experience has taught this researcher that researched writing is one of the most challenging types of writing for students. Students in composition classes are

often uninterested in the writing assignments, and have trouble getting started and staying engaged with research papers, as noted in a survey of college students' information use (Head & Eisenberg, 2010). Student engagement in researched writing was also the subject of a doctoral dissertation by Kanter (2006). Among her findings was the observation that many students find reading and integrating source material into their writing challenging. She was hardly alone in reaching this conclusion. Three decades earlier, Richard Larson, a former editor of the *Journal College Composition and Communication,* questioned the value of the traditional research paper (Larson, 1982). Larson noted that "[r]esearch can inform virtually any writing or speaking if the author wishes it to do so" (p. 813). Like Larson, Kanter suggested that substituting a different type of assignment for the traditional research paper might increase student engagement. And though the "research paper" assignment is tenacious, many scholars have agreed with them. Furthermore, Ken Macrorie (1988), who pioneered the I-Search paper as an alternative to the research paper, and many who followed him, also provided evidence that students are more engaged in writing when they are interested in the topic.

The course followed an active learning model (Bonwell & Eison, 1991) with students learning by doing, that is, by working the assignments and seeking information and help on their own. Unlike many information literacy classes, there were no mandatory lectures or demonstrations by the librarian. She assumed the role of a coach and participated in online forums when she felt she could contribute to topics under discussion or model expert thinking. Relevant content concerning research techniques and sources was made available through a course management system and the librarian was available to students for consultations in person or online. No specific feedback was

given to the students regarding their individual performance on the pretest. The subsequent phases of the course addressed each of the learning objectives and criteria. This was done intentionally to test the scaffolded curriculum and emphasis on critical thinking and self-reflection, which were also important learning objectives.

One assignment that was not scored as part of this study, but that was intended to increase active learning and engagement, and provide more practice in information skills, was the WISER Project. It had been piloted at NJIT during several prior semesters with indications of surprisingly high levels of interest and engagement as a result. For example, the first semester the Wikipedia assignment was listed on a syllabus for Technical Communication, it was the third assignment of the term. Yet students were curious about it; some began to work on it at the start of the term, even before other assignments with earlier due dates. Informal feedback from many students confirmed genuine interest in this new type of assignment. This experience echoed Macrorie's work and seemed a promising way to improve student engagement in researched writing. This thinking influenced the choice of topic for the diagnostic essay, the choice of Wikipedia as a tool of engagement, as well as the decision that students select their own topics for the WISER Project.

Writing for wikis in general, and for Wikipedia, in particular, seems to promote engagement for several reasons. An important one is the authenticity of writing for a real audience. In traditional writing assignments the students may be enervated writing for the teacher or an imagined audience, seeking to please or meet expectations rather than engage. The prospect of writing for Wikipedia seems to energize students because they are asked to write for an authentic and public audience. Researchers have also noted the

positive effects on students of moving the role of audience from the instructor to the Wikipedia community. The instructor becomes coach rather than judge. Writing for Wikipedia may also change students' understanding of the reader-writer relationship. As Kuteeva (2011) observed, writing on a wiki increased students' attention to their audience and to the mechanics of writing. Wikis are also a collaborative vehicle for the digital era, and there is a substantial literature on the uses of wikis in teaching rhetoric. The collaborative aspect of Wikipedia was initially used to pilot development of a wiki publishing environment for student work (Forte & Bruckman, 2006).

Much of the discussion of Wikipedia as a reference source concerns its quality. (Anderka,2012; Julien, 2012; Thornton-Verma, 2012). Zazzau (2009) explored the use of wiki software for library instruction as a tool for teaching critical thinking through team-based learning. Although she did not use Wikipedia, her survey of two classes revealed that none of the 44 respondents realized they could edit Wikipedia and did not view it as a social networking tool. This study used Wikipedia as a teaching tool for information literacy that promotes both cognitive and affective dimensions of learning. The literature on using Wikipedia for this purpose is limited, but the effects of wiki-writing in general are mentioned in print and online by other educators (Cummings, 2009; Forte & Bruckman, 2006; Kuteeva, 2011; O'Sullivan, 2009), many of whom can be found through the Wikipedia University Projects pages ("Wikipedia: School & University Projects," 2012), the Wikimedia outreach initiative to higher education ("Wikipedia Education Program," 2012), and a listserv moderated by Robert Cummings, "teaching-with-Wikipedia." The use of authentic research assignments has been proposed in the composition literature by Macrorie (1988) and in the library literature by Kuhlthau

et al. (2007). Other post-secondary educators have used Wikipedia to enhance content knowledge for example, in chemistry, political science, language translation, and many other subjects (Wikimedia Foundation, June 2012).

4.5.7 Wikipedia Primer

Because the instruction for this course was built around using Wikipedia, a brief introduction to some of its features is in order. Anybody with Internet access can edit Wikipedia. Setting up a login confers some privileges and provides anonymity, if desired, but is not required. The way to edit Wikipedia is simply to make changes online and, if there is disagreement from other editors, to communicate with the Wikipedia community until consensus is reached. There is a common misunderstanding that "Wikipedia" has ultimate editorial control. However, Wikipedia's content is entirely controlled by those who chose to write and edit articles—all are volunteers, called "Wikipedians." This means that even the "bots" that perform tasks such as the automatic labeling of articles for violations such as "This article does not cite any references or sources" are written by volunteer editors. No one is in charge. All the policies and procedures have been developed collaboratively by the community. There are millions of registered editors, but user activity follows the power laws: a small percentage of the users are responsible for the majority of the work (Kittur, Chi, Pendelton, Suh, & Mytkowicz, 2007). Andrew Lih (2009), a journalist and an active Wikipedian since 2003, provides good insight into how it works in his 2009 book *The Wikipedia Revolution*. The typical profile of an active Wikipedian is not unlike the student profile at NJIT—a male student in his early 20s from a developed country (Glott, Schmidt, & Ghosh, 2010). A small percentage of these editors gain administrative privileges that allow them to block users who violate

Wikipedia principles, and to delete, restore, and protect pages. These privileges are bestowed by the consensus of other "admins" to those who are active editors and have gained the trust of the community. Individual members of the online community gain or lose administrative and editorial privileges only through the consensus of other experienced editors (Wilson, 2008). Communities of interest come together in WikiProjects, where broad topics or related groups of topics are coordinated, organized, and managed through voluntary collaboration.

The project is funded by its parent organization, the Wikimedia Foundation. Wikimedia's purpose is to raise money to support the technical infrastructure and "to empower a global volunteer community to collect and develop the world's knowledge and to make it available to everyone for free, for any purpose" ("Frequently asked questions," 2012). In the Web 2.0 global world, two core values of Wikimedia, and deeply embedded in the culture of Wikipedians, are transparency and conformity with community practice. Every past edit and online conversation that occurs is always available to all. In a teaching situation, this enables the instructor to view the entire history of edits made by the student (and everyone else) if the Wikipedia user name is known. It is through the "View History" or "Discussion" tabs at the top of an article that one enters the collaborative world of Wikipedia. Although composition instructors struggle to have students put together portfolios that document their process, Wikipedia provides an open window to "process."

In addition, the overall editorial guidelines for Wikipedia provide a clear set of conventions that students should follow—as they do in their scholarly and work lives. The guiding editing principals, also developed by the volunteers, are defined as the Five

Pillars ("Wikipedia," 2012, June 14) and developed in many related articles in Wikipedia. Furthermore, Wikipedia articles provide guidance on everything from "What Wikipedia is not" and "What makes a perfect article" to how to cite sources and obtain a creative commons license and upload images ("Wikipedia: Policies & Guidelines," 2012). Thus, additional benefits of using Wikipedia as a teaching tool are that Wikipedia provides a real-time opportunity for students to understand and participate in a community of writers operating under a set of conventions that must be discovered through research and communication. Independently published, *Wikipedia: The Missing Manual* and its online updates (Broughton, 2008) provides one place where user documentation is organized and presented as a manual, but Wikimedia provides no official instruction manual and the guidelines and practices are evolving and must be discovered through searching Wikipedia and engaging with the community. Sometimes guidelines may seem contradictory or ambiguous and require research, the discovery of precedent, and judgment for interpretation, just like in the "real" world.

Finally, discussions about Wikipedia typically engender questions about neutrality and accuracy. Stephen Colbert famously mocked this aspect of Wikipedia in a satire that was shared with the students ("The Colbert Report," July 31, 2006). Despite the satire, and contrary to its reputation, studies have shown that Wikipedia is fairly accurate and becoming more so. A study by *Nature* in 2005 used peer review to compare a sample of science entries in Wikipedia and the *Encyclopedia Britannica* and found Wikipedia only slightly less accurate (Giles, 2005). The authors discovered more error in the *Britannica* than was generally imagined. A scholarly if limited study by Tom Chesney (2006) found Wikipedia's error rate to be 13%. A more detailed study in 2007

showed Wikipedia citations flourishing with a tendency towards the increasing use of the standard citation formats and agreement with patterns of citation found in scientific journals (Nielsen, 2007). A more recent study of quality in Wikipedia focused on flaws rather than accuracy, and found that 70% of the articles have at least one flaw, the most common being related to verifiability, followed by issues relating to content clean-up or lack of data, among other things (Anderka & Stein, 2012). Thus, two key Wikipedia principles—that a neutral point of view be maintained, and that information be verifiable—are frequently violated and are of special interest and an opportunity for educators. Being compelled to write for Wikipedia forces students to examine it more closely, to question the accuracy, format, completeness, neutrality, and sources of information they use, and to examine their own interpretation and communication of that information. The WISER Project also fulfilled the information literacy objectives as stated in NJIT's Information Literacy Plan (New Jersey Institute of Technology, 2009):

- recognize when information is needed;
- locate it efficiently regardless of its location, format, or medium;
- evaluate its relevance, authoritativeness, and validity;
- use it to build new knowledge;
- communicate that knowledge.

4.6 Experimental Procedures

4.6.1 Sample and Administration

A total of 274 students, enrolled in five sections offered by three professors teaching ENG352, a third-year technical communication course required for engineering

majors at NJIT, in the spring and summer semesters of 2011, were asked to participate. Table 4-4 shows the overall research design and approximate enrollment numbers. Students were asked to complete a demographic questionnaire and consent to the inclusion of their data in this study. Not all students agreed to participate. This project, protocol, questionnaire and consent form (see Appendix C) was approved as an exempt study from both the Rutgers and NJIT IRBs (See Appendix F).

Table 4-4 Experimental Design

pre	Experimental Groups	post	Approximate No. of students
x	**Treatment group**	**x**	**183**
	Control group	**x**	**91**
		TOTAL	**274**

This study sought to determine if the instructional program improved students' ability to evaluate and use appropriate sources in a persuasive argument. Information literacy was a course objective in all five sections. The diagnostic essay was administered as a pre- and posttest in the experimental sections and as a posttest to the control group. All groups received the same directions about the assignment. The students in the experimental group were not given any direct feedback on their pretest, but general feedback was part of the information literacy instruction embedded in the course content. In the experimental group, the librarian and the faculty member collaborated to create instructional materials that integrated research and writing skills into the course content. In the control group, the information literacy instruction received by students was solely from the instructor without intervention by the librarian. In all cases, the essay

instructions were administered through the course management system (Moodle) and students received credit for the assignment irrespective of their participation in this research study. All instructors were aware of the information literacy objectives for this course and had taught the course in the past.

4.7 Research Question 1 and Hypotheses

Is the brief essay a reliable and valid instrument for assessing higher-order information literacy skills?

Hypothesis 1.1: The brief essay can be reliably scored.

Consistency in scoring among several readers is an important measure of reliability when using constructed-response tests. This part of the study measured the degree to which we have confidence in the consistency of the scores given by multiple raters, that is, the degree of interrater reliability. The scoring took place over a period of several months on an irregular basis following the end of the semesters under study. The mix of scorers included writing instructors, academic librarians, education graduate students, and the researcher, all of whom were familiar with academic teaching and learning goals for information literacy. None of the raters were instructors for the course. Each essay was assigned a unique number and all identifying information was removed.

In order to achieve a high rate of consistency all scorers were trained using the same procedures. The researcher prepared an overview of the scoring procedures to orient the readers. A set of sample papers representing student work in the "accomplished," "developing," "poor," and "very poor" ranges for each of the evaluation criteria were selected by the researcher. The researcher scored each paper and made notes on key decision points to share during the session. The training session consisted of an

overview of the project, a copy of the instructions given to students, an introduction to the importance of consistency in scoring, and a brief discussion of the rubric. Scorers were asked to read and score the exemplars and make notes about their thinking, especially on difficult decision points. The researcher encouraged readers to think of the overall category first, and then consider whether it was high or low within that category. Following the reading, the researcher led a tabulation and discussion of each trait, allowing time for members of the group to discuss and understand how to apply the scoring parameters in practice.

Following the training, each essay was scored by two readers and adjudicated by a third if the scores were discrepant. The issue of assignment of essays and readers can introduce error if proper consideration is not given to the method used (Shrout & Fleiss 1979). In this case, the convenience method was used for assigning papers to readers. Assignments depended on who among the recruited readers was able to attend on a given day. Because attendance depended on the personal schedules of the readers, it is assumed that no pattern of assigning papers to readers is present, thus little error should be introduced by this method. Fifteen people were trained to score during three separate sessions and repeated calibration was handled through subsequent discussion during each scorer's first scoring session.

Administration of scoring was handled on paper. Scorers indicated their ratings, essay number, and their initials on a printed copy of the rubric for each essay. This enabled the researcher to ensure that a paper was not read twice by the same scorer. The score sheets were separated from the essays and the researcher assigned essays to second and third readers as described earlier to resolve discrepancies. The second and third

scorers did not see the scores given by previous scorers. Any score on any of the independent variables was adjudicated by a third reader if the first scores of the first two readers were discrepant. The researcher handled the administration during each session.

The individual essays were numbered randomly so that all identifying information (the student's name, date, professor, treatment group, pre or post) were masked. All information about essays and students were keyed to a database using each student's unique ID to which only the researcher had access. Several scorers dropped out of their own accord early on. Two scorers were dismissed because their ideology prevented them from adhering to the scoring norms of the group (Bar & Zussman, 2012). In cases in which there were three or more highly discrepant scores, another reader was assigned to supply a complete fresh set of scores for the paper. The two closest sets of scores were then used for the overall score calculation. Six scorers proved to be consistent and available and scored the remaining papers. A preliminary assessment of scoring reliability was made when the majority of spring 2011 essays had been scored. Following the reading of a good percentage of the spring 2011 papers (n = 190), interrater reliability was calculated and two readers were found to be in agreement on each trait over 80% of the time (See reliability results in section 5.3. Since the reliability was satisfactory, scoring was completed.

How Final Scores were Calculated

Each essay was read independently by two readers. If scores did not fall into the same category (A = 10, 9, 8; B = 7, 6, 5; C = 4, 3, 2; D = 1), the essay was read by a third reader and the discrepant trait scored. The final score for each variable was the average of two scores. There was only one adjudication, so if the first reader gave it an 8, the

second reader a 5, and the third reader a 1, there was no fourth reading, but the adjudication worked as follows: For example:

Rater 1: score = 8

Rater 2: score = 4

Adjudication was needed in this case because the scores fell into two different ranges.

Adjudicator: score = 5

In cases where no adjudication was required, it was the average of Rater 1 and Rater 2 scores. If there was adjudication, the two closest scores were averaged. Although the scores might still be discrepant, no additional adjudications were conducted. For example:

To determine which rater's scores to use for the final average, we looked at the difference between each rater's score and the adjudicator's score. For example:

Rater 1: score = 1

Rater 2: score = 2

Adjudicator: score = 3

Final score = 2,5

Because Rater 2's score is more consistent with the adjudicator's score, Rater 2 and the Adjudicator's scores would be averaged to arrive at a final score. If there is a two-category split, the higher score was used. (See Table 4-5 for sample cases.) If an essay had more than three discrepant scores, the most discrepant reader was dropped and the paper re-read by another reader. This method was used to weed out the most inconsistent readers.

Table 4-5 Scoring Examples

Rater 1 Score	Rater 2 Score	Adjudicator Score	Score 1 Adjudicated	Score 2 Adjudicated	Average
8	4	5	4	5	4.5
8	4	6	8	6	7.0
7	6	no adj as they're in the same range	7	6	6.5
8	7	7	7	7	7.0
4	5	4	4	4	4.0
5	7	no adj. required	5	7	6.0

This scoring method has been criticized because it is more time consuming than an easily scored limited-response test. The shorter than usual length of the essay was intended to minimize the time required for effective scoring.

Hypothesis 1.2: The criterion variables used will be internally consistent.

Another measure of reliability is the internal consistency of the criteria. This was calculated using Cronbach's alpha. Figure 4-3 shows the steps taken to provide an acceptable level of confidence in the two measures of reliability considered in this study and the statistical methods used to make that determination.

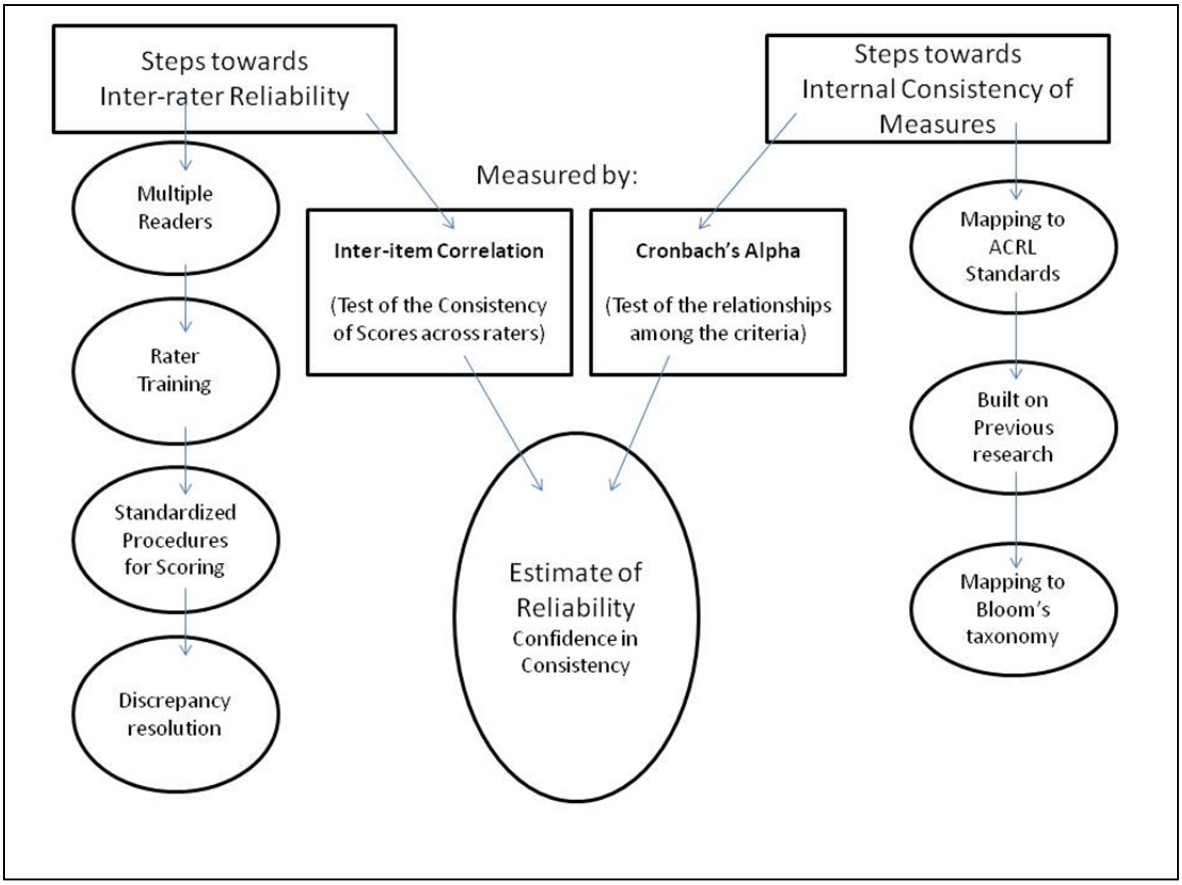


Figure 4-3 Reliability Model

Hypothesis 1.3: The brief essay is a valid instrument for assessing higher-order information literacy skills.

This part of the study evaluated issues of validity. Figure 4-4 illustrates the aspects that were considered. In a broad sense, validity refers to the degree to which the study is able to make strong links among variables (Krathwohl, 1998). The researcher followed Trochim's (2001) naming convention for the various types of validity, which he divided into two categories. The first, "translation validity," refers to face and content validity, which are not typically measured statistically, but through analytical, and therefore sometimes subjective, assessments. Translation validity is judged by whether the operationalization of the instrument and measures seem to match the concept under study. The second category is "criterion validity," which describes how well the measures relate to other measures and characteristics. It includes concurrent validity, convergent validity, discriminant validity, and predictive validity. These can be measured statistically. The student work product is the result of a complex process and that product is the evidence of it. All tasks are mediated by the judgment of the scorer and are susceptible to the same threats as other tests. Thus the construct may be under- or overrepresented and there may be unintended consequences. The statistical measurement of the relationships among variables in the study will provide evidence of validity. A summary of the types of validity addressed in this study are summarized in Figure 4-4.

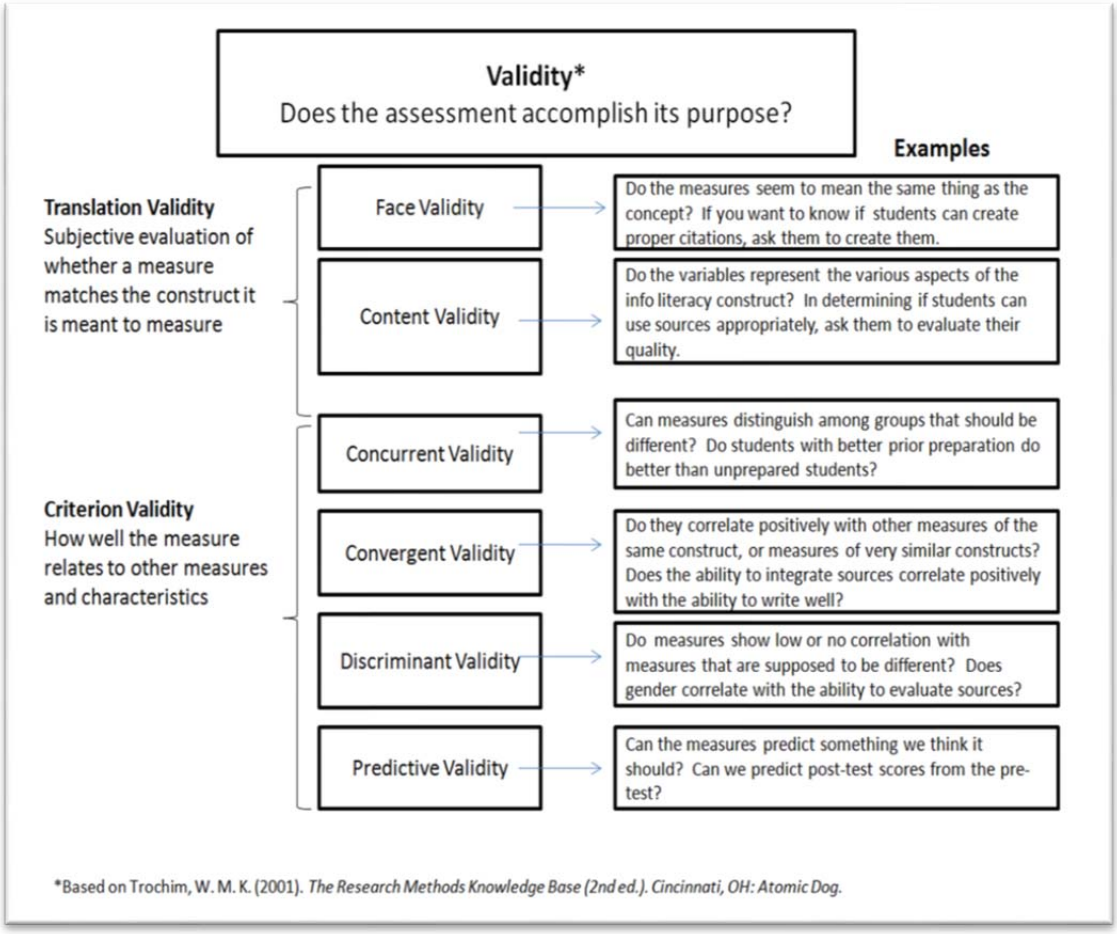


Figure 4-4. Validity Model

Face validity is judged by whether the operationalization of the instrument and measures seem to match the concept under study. Readers and instructors will be asked their opinion on this. The issue of essay length is one that may also affect face validity. Thus, the length of the essay was an important consideration. A 300-word minimum, excluding the references and reflective essay, was required, but there was no maximum length. The justification for encouraging students to write a short essay is based on previous studies that have evaluated student work such as bibliographies, term papers, and portfolios, where the quantity of student work product can make reviewing and

scoring a daunting task. In this study, the administration and scoring is made more practical by limiting the length of the essay to a page or two. It has been shown that the correlation between essay length and scores on the SAT writing test is indeed high (Kobrin, Deng, & Shaw, 2007), and the College Board has come under some criticism for this (Perelman, 2005), but this finding is not surprising. A certain length is needed to effectively develop an idea. However, unlike the essay in this study, the SAT is a timed test of writing and most writers know it takes longer to write concisely. Some of the world's best essays are brief and to the point. Furthermore, even a long abstract is widely understood as a summary of the main points of complex ideas and research. In the real world (rather than the world of standardized tests), the length of time it takes to write an essay and the length of the essay itself are choices an author makes based on the constraints of the project. Thus, it is one of the aspects taken into account in the scoring of our authentic assignment. Experience with similar assignments used for assessment studies at NJIT enabled us to set a minimum length for the text, excluding references and the reflective essay, but left the final word count up to the students because it was seen partly as a measure of how well they understood how much information was needed to adequately answer the question.

Content validity requires that the variables represent the various aspects of the information literacy construct. It was assessed by mapping the measures to the ACRL Standards, which, as noted above, are widely accepted as representing the information literacy construct. If the traits described by the rubric can be linked back to the criteria for "information literate" individuals, it strengthens the validity of the measures. Given the multitude of measures that could be mapped, why were these variables chosen? The

variables chosen and used to score the essays were based on prior work (Scharf et al., 2007) that was expanded to better elucidate the relationship between writing and research. In addition, Messick (1994) challenged us to ask if the task presented is meaningful in context and this evidence will be sought by analyzing and coding the reflective essays. This is important because the issue of motivation is one that can impose insurmountable barriers to execution (Hidi & Boscolo, 2008). Thus, creating a researchable question that is engaging for undergraduates is an important step in designing an authentic assessment, as Frey (2012) points out. In considering ***c***ontent validity we may ask: How do the criteria map to the theory and to the ACRL Standards.

Construct validity is partially addressed by looking at criterion type validity, so we may begin with concurrent validity by asking if things in the real world that should relate to each other are indeed related. Concurrent validity expects measures to distinguish among groups that should be different and was assessed by comparing subgroups of students in relation to their performance on the information literacy criteria. For example, did students who used the course materials more do better? One would expect so.

Convergent validity means measures correlate positively with other measures of the same construct, or measures of similar constructs. Theoretically, it might be assessed by comparing these measures with those on one of the standardized information literacy tests such as iSkills, SAILS, or one of the general tests of academic skills such as CLA or the ETS Proficiency Profile, but this was not feasible within the constraints of this study and must be left for future researchers. In any case, the field of information literacy testing is in its infancy and the only previous work in comparing one of them, the iSkills

test, with a similar type of constructed-response test was not successful, but it pointed up the lack of convergent validity between the two purported information literacy tests (Katz et al., 2008). The researcher's interpretation of this finding is that the standardized limited-response iSkills test, though more innovative than a multiple choice test, still cannot evaluate the type of complex constructed responses that are needed to provide evidence of information literacy. Therefore, in this study, the evaluation of convergent validity relied solely on a comparison of the performance on the information literacy criterion variables with each other and with other measures such as SATs, course grade, and GPA. Discriminant validity means measures do not correlate with variables that seem unrelated and was measured by correlations between information literacy and covariates that would not be expected to be related. Thus, in this study, diagnostic essay scores should be related to grades but not to gender. Predictive validity means a measure is able to predict something we think it should. Can we make predictions about new students? Does the pretest predict other grades, the amount of study, the posttest, or the course grade? In this study, it was measured by testing which of the variables could be predicted for new cases from pretest scores. Multivariate statistical analysis using the general linear model was used to determine if any of the significant independent variables from this study could predict performance of students taking this course in the future.

4.8 Research Question 2 and Hypotheses

Will the intervention designed by a collaborating researcher and instructor improve students' ability to use researched information effectively in their written work?

Hypothesis 2.1: There will be a statistically significant improvement in student performance on the research essay following the intervention specifically targeted at developing the component higher-order information literacy skills.

The theory underlying the idea of "improvement" is based on Vygotsky's (1978) concept of the zone of proximal development (ZPD): "the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers." (p. 86). The diagnostic essay seeks to measure the baseline of each student entering the course in the pretest, and the distance traveled by each one within his or her ZPD for the abilities related to information literacy in the posttest.

An experimental method was used to test the hypothesis that student work would improve following the intervention. A pretest and a posttest were administered to the students in the treatment group. Because different methods of analysis can result in different interpretations, several methods of data analysis were tried. First, a paired sample *t*-test and a repeated measures ANOVA were used to determine if there was a significant change in information literacy performance by the end of the semester. This compared the mean change from pre- to posttest to answer the question in a simple way without taking covariates into account. The gain scores, the difference between the scores on the pre- and posttests, were also calculated to determine if students improved significantly following the treatment and an independent *t*-test was performed using the gain scores. Next, an analysis of covariance (ANCOVA) was performed to answer a slightly more complex question. What effect did the treatment have on the posttest that

was not predicted by the pretest? In this case, the group was the principal independent variable, with the posttest score as the dependent variable. (See Figure 4-5.) To test for a significant improvement in student performance on the diagnostic essay following instruction, intended to develop the component higher-order information literacy skills, a quasi-experimental research design was used. This was chosen because creating perfect experimental conditions using randomized test and comparison groups is rarely possible in an academic setting.

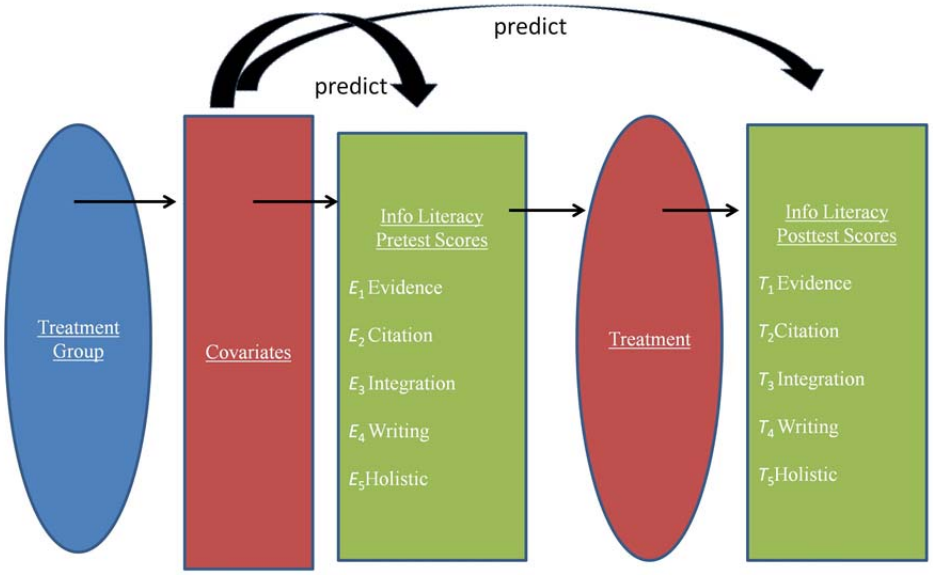


Figure 4-5. Design to Compare Performance Before and After Instruction.

Hypothesis 2.2. Students in the treatment group will perform significantly better than those in the control group.

In this study, students in other sections, taught by other instructors, were given only the posttest during the last few weeks of the course in order to compare students who did not go through the intervention and to create a performance benchmark for

students not in the treatment group. The students in the control group took the same course, with the same objectives, as the treated students, but it was taught by different instructors, without the collaboration of the librarian or the information literacy intervention described previously. Furthermore, the control group should provide scores for a similar population of students outside the study as well as additional data for testing the reliability and validity of the instrument. Figure 4-6 depicts the relationships of the variables related to testing the effect of the scaffolded instruction on higher-order information literacy skills on the treatment and control groups. Lacking ideal experimental conditions, it was not possible for the students in the control group to take the pretest as well, but the effects of the pretest were removed statistically from the result. Such a study may be carried out at a later time when pre- and posttests can be administered to both treatment and control groups.

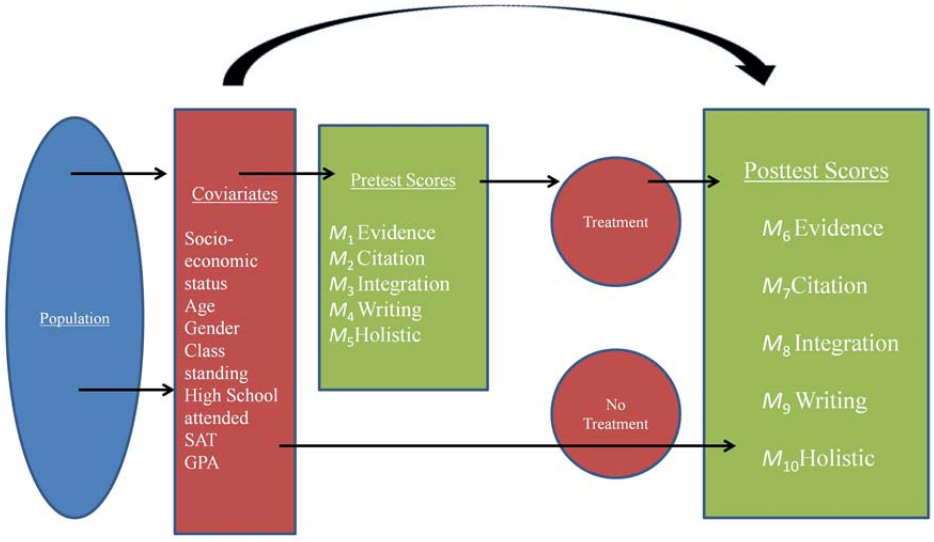


Figure 4-6. Design for Comparison Study of Treatment and Untreated Students

4.9 Research Question 3 and Hypotheses

Which prior conditions have a significant effect on a student's information literacy performance?

Hypothesis 3.1: The pretest accounts for a portion of the variance in scores between the pre- and posttest.

Several covariates representing various prior conditions were tested to determine if they account for any of the variance in pre- and posttest scores. The relationships of the variables to admissions tests (the SATs), grade point average, and gender were also calculated to see if any other important variables had a significant or confounding effect. Statistical analysis using the general linear model was used to account for the pretest bias and pretest and posttest scores between treatment groups. Figure 4-7 shows the relationships of the variables related to testing the effect of prior conditions on information literacy.

Hypothesis 3.2. Socioeconomic status will show a significant correlation with information literacy scores.

As discussed in the literature review, researchers have often shown strong correlations between academic performance and indicators of socioeconomic status. To test the hypothesis that SES is a marker for information literacy, a variable that represents SES, was correlated with the scores on the diagnostic essay to determine if any of these were markers for information literacy. In his meta-analytic review of the research studies on SES and academic achievement, Sirin (2005) discusses the various SES measures that have been used. He cites the three most widely used measures of SES, parental income, parental education, and parental occupation, and notes that parental education is one of

the most stable indicators (p. 419). SES indexes of occupation have shown that parental education is also highly correlated with income in the United States, so for this and other reasons it has been suggested that this indicator alone is a good measure (Hauser & Warren, 1996). In addition, this measure is part of the student record and is easily obtained.

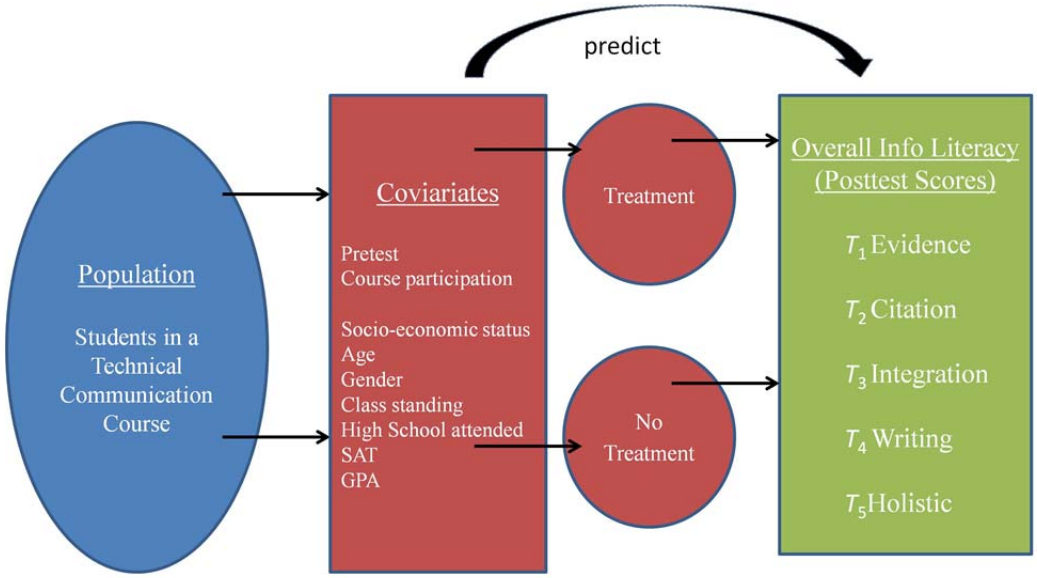


Figure 4-7. Design for Testing the Effect of Prior Conditions on Information Literacy

Demographic data (age, academic standing, parents' education level) and consent to participate in the study was collected via an online survey tool. Additional data on academic performance was also collected so that the effects of confounding variables could be isolated. SAT, GPA, and final course grades were harvested from the student information system. Grades on individual assignments were retrieved from Moodle, the learning management system.

4.10 Research Question 4 and Hypotheses

What can we learn from how students use course materials and understand their own learning?

Hypothesis 4.1. Frequency of access of materials correlates with performance.

This portion of the study relied on data collected automatically through Moodle. This data set consisted of quantitative data on how frequently each student accessed relevant instructional materials. This data was used to examine usage and the potential value of the materials. Additional analyses were performed to see if the degree or type of activity on Moodle throughout the semester predicted essay performance. (See Figure 4-8)

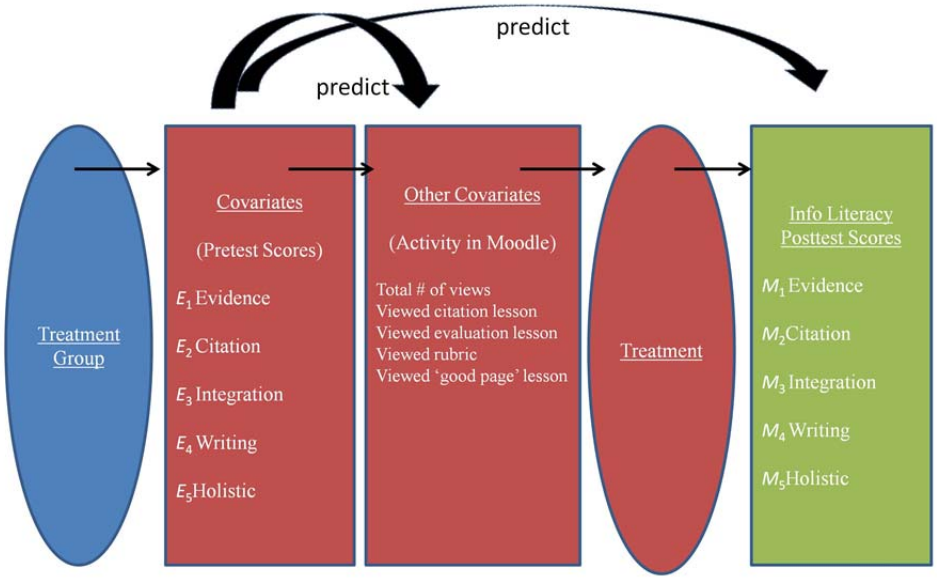


Figure 4-8. Moodle Activity Predicts Performance.

Hypothesis 4.2. Students' understanding of their own research/learning process correlates with performance.

In order to explore this part of the question, another set of data was obtained by using content analysis to understand the implications of the reflective statements students were asked to write. Students were given the following instructions:

> For this part, you are asked to explain the thinking and process you used to complete the first part of the task. Please explicitly describe how you first engaged with the topic, how you found material to support your points used in the essay, and how you determined these sources to be reliable, trustworthy, and of quality for the essay. This reflective statement can be as long as you feel is needed to fully describe what is asked.

Content analysis may be both quantitative and qualitative and both methods were used here. First, a conceptual analysis was conducted by the researcher on a subset of the statements to develop a framework of categories. Then, the entire set of reflective statements was analyzed by tallying the occurrences of statements that fit these categories. The statements were used to address the following questions. Is there evidence that students . . .

- read the instructions?
- sought out the full text of the suggested sources?
- conducted "authentic" research?
- knew how to use the library tools and sources?
- considered alternative viewpoints when conducting the research?
- were conscious of their selection criteria and the quality of sources?

Hypothesis 4.3. Essay length correlates with performance.

It has been shown that longer essay length correlates with higher scores on a timed standardized test of writing (Burstein, Chodorow, & Leacock, 2004). In this study, a minimum length of 300 words was required, but no maximum was specified. In this study, correlation was used to determine if the "pretest essay word count" and "pretest reflective statement word count" correlated with information literacy achievement.

4.11 Prediction

Can we predict posttest scores for the next semester from the pretest? If students score high on these measures, the diagnostic should be able to predict new cases, as shown in Figure 4-9. Can pretest scores be predicted from word count or another factor?

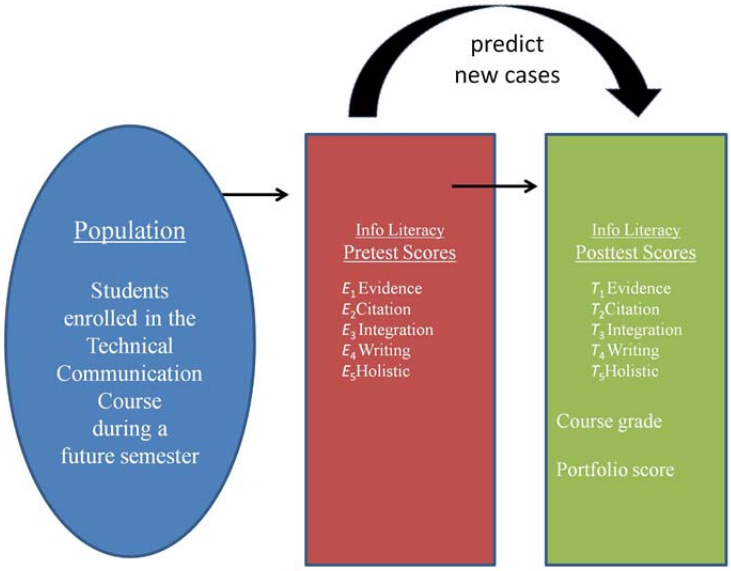


Figure 4-9. Design Demonstrating New Case Prediction

Table 4-6 lists the variables under consideration. The unit of analysis is the student. Table 4-7 is a summary of the research questions, hypotheses, relevant variables, methods of analysis, and expected findings.

Table 4-6. Summary of the Variables.

Independent Variables (IV)	**Operationalization**
Preconditions	
IV1 - Socioeconomic status and education	self-reported on survey (or SIS)
IV2 - High School GPA	self-reported on survey (or SIS)
IV3 – SAT	self-reported on survey (or SIS)
IV4 – College GPA (precondition)	self-reported on survey (or SIS)
IV5 - Gender (precondition)	self-reported on survey (or SIS)
IV6 – Age (precondition)	self-reported on survey (or SIS)
IV7 – Pretest (as covariate in treatment group)	index of total diagnostic essay score
IV7.1 – Pretest Evidence of Research	diagnostic essay score
IV7.2 – Pretest Mechanics of Citation	diagnostic essay score
IV7.3 – Pretest Integration and Ideas	diagnostic essay score
IV7.4 – Pretest Writing	diagnostic essay score
IV7.5 – Pretest Holistic Information Literacy	diagnostic essay score
Instruction	
IV8 - Treatment/Instruction (instructional strategy)	collaborative instruction or control
Assessment	
IV9 - Student behavior (assessment)	activity reports from course mgmt system showing how much effort student expended
IV10 – Degree of self-reflection (assessment)	reflective essay scores
Dependent Variables (DV)	**Operationalization**
DV1 – Pretest measure of information literacy	index of total diagnostic essay score
DV1.1 – Pretest Evidence of Research	diagnostic essay score
DV1.2 – Pretest Mechanics of Citation	diagnostic essay score
DV1.3 – Pretest Integration and Ideas	diagnostic essay score
DV1.4 – Pretest Writing	diagnostic essay score
DV1.5 – Pretest Holistic Information Literacy	diagnostic essay score
DV2 – Posttest measure of information literacy	Index of diagnostic essay score
DV2.1 – Posttest Evidence of Research	diagnostic essay score
DV2.2 – Posttest Mechanics of Citation	diagnostic essay score
DV2.3 – Posttest Integration and Ideas	diagnostic essay score
DV2.4 – Posttest Writing	diagnostic essay score
DV2.5 – Posttest Holistic Information Literacy	diagnostic essay score
DV3 – Course grade	Prof's course grade

Table 4-7. Summary of Proposed Data Analysis Method

Research Question	Variables	Applicable Statistical Method	Expected Finding
RQ1. Is the brief essay a reliable and valid instrument for assessing higher-order information literacy skills?			
H 1.1. The brief essay can be reliably scored.	DV1.1-DV1.5 DV2.1-DV2.5	Inter-item correlation	significant correlations
H 1.2. The criterion variables used will be internally consistent.	DV1.1-DV1.5 DV2.1-DV2.5	Cronbach's α Pearson's r Intra-class correlation coefficient	▪ High coefficients ▪ high r's ▪ high rate of consistency
H 1.3. The brief essay is a valid instrument for assessing higher-order information literacy skills.	DV1.1-DV1.5 DV2.1-DV2.5 IV3-4	Correlation ANOVA	Correlations among similar measures of IL, yet significant differences in scoring
RQ2. Will the intervention designed by the collaborating researcher and instructor improve students' ability to use researched information effectively in their written work?			
H 2.1. There will be a statistically significant improvement in student performance on the research essay following the intervention specifically targeted at developing the component higher-order information literacy skills.	IV7 → DV2	▪ Repeated Measures ANOVA ▪ Compare Pretest and GAIN Scores (Paired sample t-test)	▪ Positive gains will be significant. ▪ Some traits will improve more than others on avg. ▪ sig diff. between GAIN score means and pretest means
H2.2. Students in the treatment group will perform significantly better than those in the control group.	IV8 → DV2	▪ Independent sample t-test for group means & low/high skills ▪ One-way ANOVA	▪ sig positive difference between groups means at high confidence interval

Research Question	Variables	Applicable Statistical Method	Expected Finding
RQ3. Which prior conditions had a significant effect on a student's information literacy performance?			
H 3.1. The pretest accounts for a portion of the variance in scores between the pre- and posttest.	IV1-6, 8,9,10 DV1-3	ANCOVA to eliminate effects of pretest	▪ significant effect sizes even when effects of pretest removed
H 3.2. SES and other factors will show a significant correlation with information literacy scores. How much of the variability do they explain?	IV1-6, 8,9,10 DV1-3	Factorial ANOVA Separate models for each DV	▪ High correlations between some of the predictor variables and the DV's. ▪Some factors will have a significant effect size (eta squared). ▪ Information from each trait unique (not too much overlap).
RQ4. What can we learn from how students use course materials and understand their own learning?			
H4.1. Frequency of access of materials correlates with performance.	IV 9 DV1-3	ANOVA	Some factors will have a significant effect size (eta squared). Significant relationship between Moodle accesses and info lit posttest scores.
H4.2. Student understanding of their own research/learning process correlates to their performance.	IV10 DV1-3	ANOVA	Significant relationship between self-knowledge variables from reflective statement and info lit scores
H.4.3. Essay length correlates with performance.	IV11 DV1-3	ANOVA	Significant relationship between essay length and info lit scores

4.12 Limitations

Random sampling was not feasible with the sample size and experimental conditions, so a quasi-experimental design was employed. The possibility of interaction among participants outside of class represented a threat to validity, so the students were specifically told to work independently. The wording of the prompts, inclusion of research hints, as well as the varying difficulty of both subject matter and availability of resources, is a concern, because these differences might change the level of difficulty or challenge some students more than others, depending on their prior knowledge. A future study should examine if a similar type of essay on another subject can be used effectively for the same purposes. It is likely that a sequence of several studies will be necessary in the future to test alternative prompts as researchers at ETS have done for writing assessment (Bridgeman, Trapani, & Bivens-Tatum, 2011). In addition, the rubric may benefit from repeated calibration. Despite these challenges, however, this effort was undertaken because we recognized an adaptable practical tool is needed for everyday assessment, though the routine application of rigorous experimental methods is only occasionally feasible for teaching faculty.

CHAPTER 5: RESULTS

5.1 Participants and Demographics

A total of 274 undergraduate students were enrolled in three sections of Technical Communication (each section taught by a different professor) in the spring and summer semesters of 2011.

Table 5-1 provides a breakdown showing how many students agreed to participate in the study and subsequently submitted diagnostic essays. Eighty-five percent of the students in the treatment group agreed to participate; of those, 65% submitted both the pre- and post-essay, whereas 75% submitted at least the post-essay. In the control group, 75% agreed to participate and 75% submitted the required post-essay. Thus, the yields from each group were similar enough not to cause concern.

Table 5-1. Numbers of Participants by Treatment Group: Spring and Summer 2011

Course Sections	**Students and Essays**				
	Enrolled students	*Agreed*	*Pairs of Essays*	*Pretests*	*Posttests*
Spring 2011 Treatment (3 sections)	129	112	83	85	95
Spring 2011 Control (3 sections)	51	37	N/A	N/A	27
Summer 2011 Treatment (2 sections)	54	44	25	25	26
Summer 2011 Control (3 sections)	40	31	N/A	N/A	26
Total Treatment	183	156	108	85	121
Total Control	91	68	N/A	N/A	53
Totals	274	224	108	85	174

Totals	*Number of students and essays used in the study*
Enrolled	*Number of students enrolled in each section at the start of the semester*
Agreed	*Number of students who agreed to participate in the study*
Pairs	*Of those who agreed, the number of students who submitted both a pre and a post essay*
Pretest only	*Of those who agreed, the number of students who submitted only the pretest*
All Pretests	*Of those who agreed, the number includes students who submitted pairs and pretests only*
Posttest only	*Of those who agreed, the number of students who submitted only the posttest*
All Posttests	*Of those who agreed, the number of students who submitted pretests and those who did not*

Demographics

The population of students taking Technical Communication used in this study generally reflected the demographics of the overall population at NJIT.

Gender

The majority, approximately 80% of the students overall and within each treatment group were male. The percentages represented in Table 5-2 are similar to the overall ratio of male to female students at NJIT and in engineering schools in general. As expected, gender did not have a significant impact on information literacy variables.

Table 5-2. Gender Distribution

Treatment group		male		female	
	N	n	%	n	%
control	53	41	77%	12	23%
treatment	106	87	82%	19	18%
total	159	128	81%	31	19%

Class Standing

Most of the students in the study were upper-division undergraduates. To be classified as a sophomore, a student must have completed 28 credits, a junior will have completed 56 credits, and a senior must have at least 90 credits. The control group was about evenly divided between juniors and seniors plus four sophomores. The treatment group had 18% more seniors than juniors than the control group, and just two sophomores. Percentages by class standing for treatment and control groups are shown in Table 5-3. It is possible that the slight skew towards higher class standing in the treatment group may have exaggerated the positive results.

Table 5-3. Class Standing Demographics by Treatment Group

treatment group	class standing	Mean	Std. Deviation	N	Percent
control	sophomores	4.875	2.5941	4	7.55%
	juniors	3.420	1.7951	25	47.17%
	seniors	4.167	2.1451	24	45.28%
	Total	3.868	2.0314	53	100.00%
treatment	sophomores	8.500	.7071	2	1.90%
	juniors	4.833	2.4636	42	40.00%
	seniors	5.918	2.2327	61	58.10%
	Total	5.533	2.3932	105	100.00%
Total	sophomores	6.083	2.7644	6	3.80%
	juniors	4.306	2.3273	67	42.41%
	seniors	5.424	2.3344	85	53.80%
	Total	4.975	2.4047	158	100.00%

Age

Most students were in their 20s, but each cohort had some older students. The distribution of students by age is represented in Figure 5-1. Age was not a significant factor in any of the posttest variables.

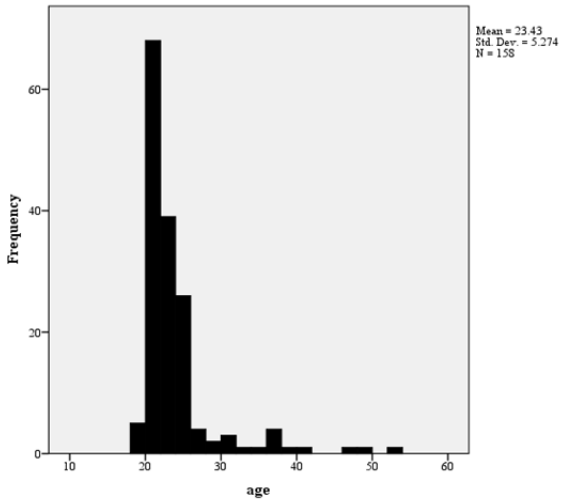


Figure 5-1. Overall Distribution of Students by Age

Majors

About 80% of the participants in the study were majoring in science, engineering, or computer science. This distribution is more heavily weighted towards the science and technology disciplines than the population as a whole because the course is required of all technical majors. About half of the undergraduates at NJIT are STEM majors and they are typically thought to be the stronger students. Table 5-4 indicates the breakdown of students in the study by major and compares performance on Evidence of Research for both treatment and control groups. Overall, the STEM majors performed slightly better, but the number of non-STEM majors in the study was small, as was the ratio of non-STEM to STEM majors. A larger sample might produce different results.

Table 5-4. Mean Information Literacy Scores for Evidence of Research Posttest by Major

Treatment Group	Major	Mean	Std. Deviation	N	Percent
control	1	3.793	2.191	41	80.39%
	2	3.85	1.375	10	19.61%
	Total	3.804	2.045	51	100.00%
treatment	1	5.679	2.369	84	79.25%
	2	5.091	2.486	22	20.75%
	Total	5.557	2.394	106	100.00%
Total	1	5.06	2.469	125	79.62%
	2	4.703	2.254	32	20.38%
	Total	4.987	2.424	157	100.00%

1=STEM major; 2=non-STEM major

Transfer Students

About 60% of the students in this study indicated they were transfer students. This differs from the overall undergraduate population for the year 2008 when many of the students in the study were first admitted to NJIT. As reported by NJIT to the National Center for Education Statistics, 60% of new undergraduates were first-time freshmen, whereas 30% were new transfers and 11% were readmissions (i.e., students who already finished at least one semester at NJIT, but then took off a full semester or longer.) No significant relationship was found between any of the pre- and posttest scores and transfer status, with or without accounting for SES. Table 5-5 indicates the breakdown by transfer status. Overall, the transfer students performed better in both groups. It is possible that the students who transferred to NJIT from other institutions did so because of the STEM program emphasis. In addition, because these transfers often come to NJIT as upper-division students, they may have been more motivated and directed than students who came to NJIT as first-time, full-time freshmen.

Table 5-5. Mean Information Literacy Scores for Evidence of Research Posttest by Transfer Status

treatment group	transfer student	Mean	Std. Deviation	N	Percent
control	non-transfer student	3.583	2.2833	18	35.29%
	transfer student	3.924	1.9289	33	64.71%
	Total	3.804	2.0447	51	100.00%
treatment	non-transfer student	4.869	2.0573	42	40.38%
	transfer student	5.984	2.5317	62	59.62%
	Total	5.534	2.4048	104	100.00%
Total	non-transfer student	4.483	2.1901	60	38.71%
	transfer student	5.268	2.5294	95	61.29%
	Total	4.965	2.4269	155	100.00%

5.2 Findings. Research Question 1. Hypothesis 1.1. Interrater Reliability

Hypothesis 1.1. The brief essay can be reliably scored.

The first research question concerns the reliability of the instrument and the process for evaluating the student work used to carry out the study. The first part of the question relates to reliability in scoring. Several statistical tests were conducted to demonstrate that the procedures designed to create scoring consistency among readers were effective. If such statistical tests indicate low reliability, the results of subsequent inferential tests would be questionable, so several measures of reliability were considered. Before proceeding to score the entire set of student essays collected, a preliminary study was made of the majority of papers submitted in spring 2011 (n = 190) to determine if the training procedures were producing reliable ratings. Results showed over 80% agreement among scorers (see Table 5-6). Interitem correlation, an additional method for confirming the reliability of the scoring was used to compare the scores of different readers on the same item.

Table 5-6. Preliminary Reliability Calculation Spring 2011: Percent Agreement

Trait	Percent agreement (n = 190)
Evidence of Research	88.9
Citation	85.3
Integration	83.6
Writing	83.6
Holistic	80.0

The interitem correlation matrix correlates the scores of Readers 1 and 2 for each trait. High correlations were found between the scores of the two readers for each of the 10 variables, indicating a high level of agreement among readers. This enabled scoring to proceed with confidence that the rating was being carried out at an acceptable level of interrater reliability. Table 5-7 shows the details of the reliability calculations for all the criteria on a large subset (n = 190) of papers scored from spring 2011.

Table 5-7. Preliminary Reliability Calculation Spring 2011-Interitem Correlation Matrix, Spring 2011

	c1s1	**c1s2**	**c2s1**	**c2s2**	**c3s1**	**c3s2**	**c4s1**	**c4s2**	**c5s1**	**c5s2**
c1s1	1.000	.876	.595	.509	.763	.701	.692	.634	.684	.646
c1s2		1.000	.589	.570	.749	.745	.694	.678	.685	.711
c2s1			1.000	.833	.663	.590	.604	.557	.567	.530
c2s2				1.000	.591	.594	.547	.547	.528	.570
c3s1					1.000	.839	.805	.719	.829	.769
c3s2						1.000	.700	.798	.765	.839
c4s1							1.000	.759	.850	.776
c4s2								1.000	.782	.866
c5s1									1.000	.818
c5s2										1.000

(n=190) c=criterion; s=score
c1=evidence of research, s1=score 1; s2=score 2,
c2=citation
c3=integration
c4=writing
c5=holistic

Following completion of the scoring, interrater reliability was measured by Cronbach's α. Table 5-8 lists the descriptive statistics for each scorer, and shows the means and standard deviations for each item to be close. The final result showed a high rate of consistency for the 10 scores (pre- and posttest criteria), among eight scorers and 283 individual essays scored (Cronbach's α = .959).

Table 5-8. Descriptive Statistics and Scoring Reliability for Spring and Summer 2011 Combined

	Mean	Std. Deviation	N
c1s1	4.59	2.480	283
c1s2	4.60	2.373	283
c2s1	5.20	2.846	283
c2s2	5.26	2.818	283
c3s1	4.19	2.294	283
c3s2	4.19	2.315	283
c4s1	4.43	2.078	283
c4s2	4.46	2.090	283
c5s1	3.88	2.233	283
c5s2	3.94	2.238	283

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Interitem Correlations	.716	.555	.943	.388	1.698	.013	10

(n = 283)

The interitem correlation matrix, which correlates readers' scores was rerun on the full data set when all the student essays had been scored. Results confirmed that the scores of Readers 1 and 2 were highly correlated for each trait, demonstrating consistency among different scorers. (See Table 5-9).

Table 5-9. Interitem Correlation Matrix for Interrater Reliability for Spring and Summer 2011 Combined

	c1adj1	c1adj2	c2adj1	c2adj2	c3adj1	c3adj2	c4adj1	c4adj2	c5adj1	c5adj2
c1adj1	1.000	.915	.590	.561	.747	.705	.653	.664	.696	.668
c1adj2		1.000	.555	.560	.717	.696	.662	.677	.676	.675
c2adj1			1.000	.943	.692	.635	.620	.576	.625	.583
c2adj2				1.000	.661	.621	.603	.570	.602	.576
c3adj1					1.000	.913	.810	.781	.844	.812
c3adj2						1.000	.775	.808	.834	.857
c4adj1							1.000	.874	.836	.781
c4adj2								1.000	.810	.839
c5adj1									1.000	.916
c5adj2										1.000

c1=evidence of research, adj1=adjudicated score 1; adj2=adjudicated score 2; *adj indicates that the scores used here were the ones used to make the final score calculation*
c2=citation
c3=integration
c4=writing
c5=holistic

The Corrected Item Total Correlation show each score correlated with the sum of the other scores for each criterion if the item were deleted. Because the result indicates each item is highly correlated with all the others, each individual score is a good component of the summed total score for each trait (see Table 5-10). If an item was deleted it would raise Cronbach's α, as shown in the last column, but it was already quite high and thus this supports using all the items and scores.

Table 5-10. Corrected Item-Total Correlation

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
c1adj1	40.15	338.193	.792	.865	.955
c1adj2	40.15	342.191	.783	.857	.956
c2adj1	39.55	330.582	.753	.904	.958
c2adj2	39.48	332.953	.736	.896	.959
c3adj1	40.55	335.567	.902	.886	.951
c3adj2	40.55	336.560	.879	.879	.952
c4adj1	40.31	346.848	.846	.839	.954
c4adj2	40.29	346.772	.841	.840	.954
c5adj1	40.87	339.485	.877	.893	.952
c5adj2	40.81	340.845	.856	.892	.953

Intraclass correlation, a slightly different method of calculating correlations among variables, was also carried out with similar results. The correlations were significant and high for single measures (ICC = .699) and for average measures (ICC = .959) as shown in.Table 5-11

Table 5-11. Intraclass Correlation

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
44.75	416.609	20.411	10

Intraclass Correlation Coefficient

		95% Confidence Interval		F Test with True Value 0			
	Intraclass Correlation[a]	Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.699[b]	.661	.737	24.245	282	2538	.000
Average Measures	.959[c]	.951	.966	24.245	282	2538	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In summary, interrater reliability, that is, consistency in scoring among different scorers, is high. This indicates that well-trained readers using clearly defined criteria can come together in agreement and we can have confidence that the scores used in subsequent analyses in this study are meaningful.

5.3 Findings. Research Question 1. Hypothesis 1.2: Internal Consistency

Hypothesis 1.2. The criterion variables used will be internally consistent.

Another measure of reliability is the internal consistency of the criteria, the degree to which the items measure the same thing. If they measure exactly the same thing, all the measures are not needed. If they have high positive correlations, it indicates that the traits measure elements of the overall construct (i.e., information literacy). The scores for students who took both pre- and posttests were used for this purpose. Table 5-12 shows the means and standard deviations for each of the variables.

Table 5-12. Descriptive Statistics: Pretest and Posttest (n = 106)

Pre- and Posttest variables	*Mean*	*Standard Deviation*
Evidence pretest	*3.76*	*2.09*
Citation pretest	*4.76*	*2.64*
Integration pretest	*3.59*	*1.95*
Writing pretest	*4.04*	*1.76*
Holistic pretest	*3.45*	*1.74*
Evidence posttest	*5.54*	*2.38*
Citation posttest	*6.15*	*2.63*
Integration posttest	*5.20*	*2.45*
Writing posttest	*5.06*	*2.16*
Holistic posttest	*4.84*	*2.36*

The scores on the pre and posttests were highly correlated for Evidence of Research $r(105) = .55$, $p \leq .0001$; Citation $r(105)=.59$, $p \leq .0001$; Integration $r(105) = .55$, $p \leq = .0001$; Writing $r(105) = .55$, $p \leq = .0001$; Holistic $r(105) = .58$, $p \leq = .0001$. The correlations of over 55% for each criterion between the pre- and posttest scores show that, to a great extent, students who did well on the one did well on the other and vice-versa. This confirms much anecdotal experience that although there may be overall

improvement in a class due to specific instruction, a strong relationship exists between the quality of work a student produces upon entering a course with work produced by that student at the end of the term. Thus, the rubric has good internal consistency. In summary, the internal consistency of the criteria is high, so we can have confidence that the criteria are related in a meaningful way. This is of particular interest because this finding supports the theory that information literacy and writing can be well represented by traits that evaluate these traits both separately and in combination.

Pearson's r indicates the linear relationship among the variables. All the dependent variables for the pretest and the posttest were found to be strongly and positively correlated. The scores on the pre- and posttests were highly correlated for Evidence of Research $r(105) = .55$, $p \leq .0001$; Citation $r(105) = .59$, $p \leq .0001$; Integration $r(105) = .55$, $p \leq .0001$; Writing $r(105) = .55$, $p \leq .0001$; and Holistic $r(105) = .58$, $p \leq .0001$. The correlations of over 55% for each criterion between the pre- and posttest scores show that, to a great extent, students who did well on the one did well on the other and vice versa (see Table 5-13). This confirms much anecdotal experience that although there may be overall improvement in a class due to specific instruction, a strong relationship exists between the quality of work a student produces upon entering a course with work produced by that student at the end of the term. This indicates the rubric has good internal consistency. In summary, the internal consistency of the criteria is high, so we can have confidence that the criteria are related in a meaningful way.

Table 5-13. Correlations: Pre- and Posttest for Treatment Group

		E1	E2	E3	E4	E5	T1	T2	T3	T4	T5
E1eviadjtotal	Pearson Corr	1.000	.493**	.707**	.673**	.690**	.552**	.370**	.473**	.509**	.503**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.000
E2citadjtotal	Pearson Corr		1.000	.641**	.547**	.611**	.399**	.597**	.428**	.476**	.419**
	Sig. (2-tailed)			.000	.000	.000	.000	.000	.000	.000	.000
E3intadjtotal	Pearson Corr			1.000	.795**	.876**	.515**	.496**	.562**	.615**	.574**
	Sig. (2-tailed)				.000	.000	.000	.000	.000	.000	.000
E4writadjtotal	Pearson Corr				1.000	.875**	.441**	.366**	.482**	.552**	.530**
	Sig. (2-tailed)					.000	.000	.000	.000	.000	.000
E5holadjtotal	Pearson Corr					1.000	.460**	.471**	.533**	.594**	.584**
	Sig. (2-tailed)						.000	.000	.000	.000	.000
T1eviadjtotal	Pearson Corr						1.000	.681**	.765**	.736**	.746**
	Sig. (2-tailed)							.000	.000	.000	.000
T2citadjtotal	Pearson Corr							1.000	.699**	.716**	.643**
	Sig. (2-tailed)								.000	.000	.000
T3intadjtotal	Pearson Corr								1.000	.913**	.903**
	Sig. (2-tailed)									.000	.000
T4writadjtotal	Pearson Corr									1.000	.928**
	Sig. (2-tailed)										.000
T5holadjtotal	Pearson Corr										1.000
	Sig. (2-tailed)										

**. Correlation is significant at the 0.01 level (2-tailed).

n=106

5.4 Findings. Research Question 1. Hypothesis 1.3. Validity

The brief essay is a valid instrument for assessing higher-order information literacy skills.

Face validity is judged by whether the instrument and measures seem to match the concept under study. When asked about this, 18 readers (including instructors, and librarians) found the assignment and the criteria made sense as tools to evoke student work that exhibited the criteria demonstrating information literacy. As expected a few (three people) expressed concern about the length of the essay. In fact, essay length did correlate with student performance. To establish content validity the information literacy criteria were mapped to the Standards as well as to Bloom's taxonomy of educational objectives (see chapter 4). These mappings are strong indications of content validity, and that the criteria are valid measures of the information literacy construct as well as of widely used concepts in learning theory and practice.

Different Instructors

Three different instructors taught the students in the study, a condition that could represent a threat to internal validity. However, course grades for both treatment and control groups showed similar grade distributions (See Figure 5-2), indicating that the three instructors graded with consistency across the three sections. In addition, an analysis of variance showed that the difference between Instructors' grades (Course grade) and the judgments of independent scorers (as represented by the Evidence posttest) were significant, $F(4,159) = 6.191$, $p \leq .0001$ (see, Table 5-14, Table 5-15, and Figure 5-3).

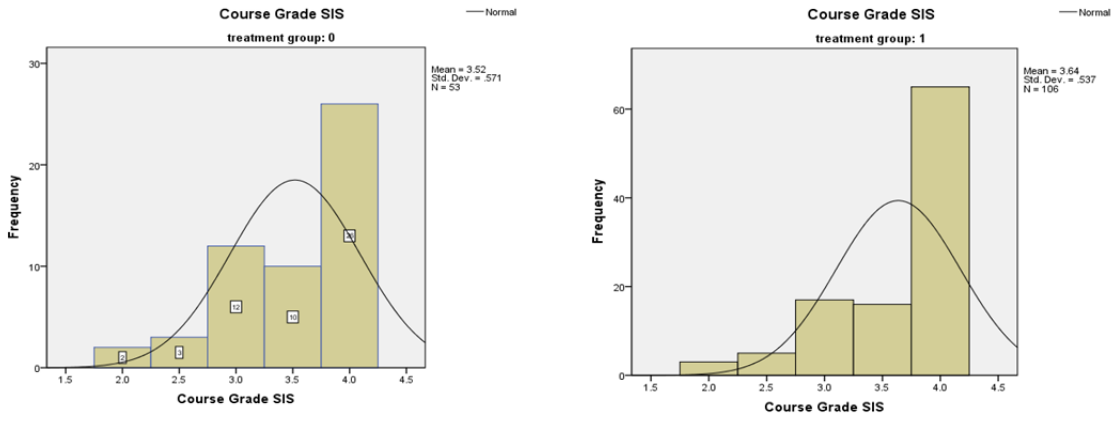


Figure 5-2. Distributions of Course Grades

Table 5-14. Relationship of Course Grade and Evidence Posttest Scores

Course Grade	Mean	Std. Deviation	N
2.0	2.200	1.3038	5
2.5	3.000	.7559	8
3.0	4.466	2.3790	29
3.5	4.423	1.8957	26
4.0	5.654	2.4286	91
Total	4.994	2.4091	159

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	127.036[a]	4	31.759	6.191	.000	.139
Intercept	953.124	1	953.124	185.809	.000	.547
Course Grade	127.036	4	31.759	6.191	.000	.139
Error	789.958	154	5.130			
Total	4882.000	159				
Corrected Total	916.994	158				

Course Grade	Mean	Std. Deviation	N
2.0	2.200	1.3038	5
2.5	3.000	.7559	8
3.0	4.466	2.3790	29
3.5	4.423	1.8957	26
4.0	5.654	2.4286	91

a. R Squared = .139 (Adjusted R Squared = .116)
$p \leq .05$

Table 5-15. Course Grade and Evidence-Posttest Scores by Treatment Group

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	215.597[a]	9	23.955	5.089	.000	.235
Intercept	832.788	1	832.788	176.912	.000	.543
CourseGrade	86.158	4	21.539	4.576	.002	.109
TREATMT	19.020	1	19.020	4.040	.046	.026
CourseGrade * TREATMT	7.262	4	1.815	.386	.819	.010
Error	701.396	149	4.707			
Total	4882.000	159				
Corrected Total	916.994	158				

a. R Squared = .235 (Adjusted R Squared = .189)

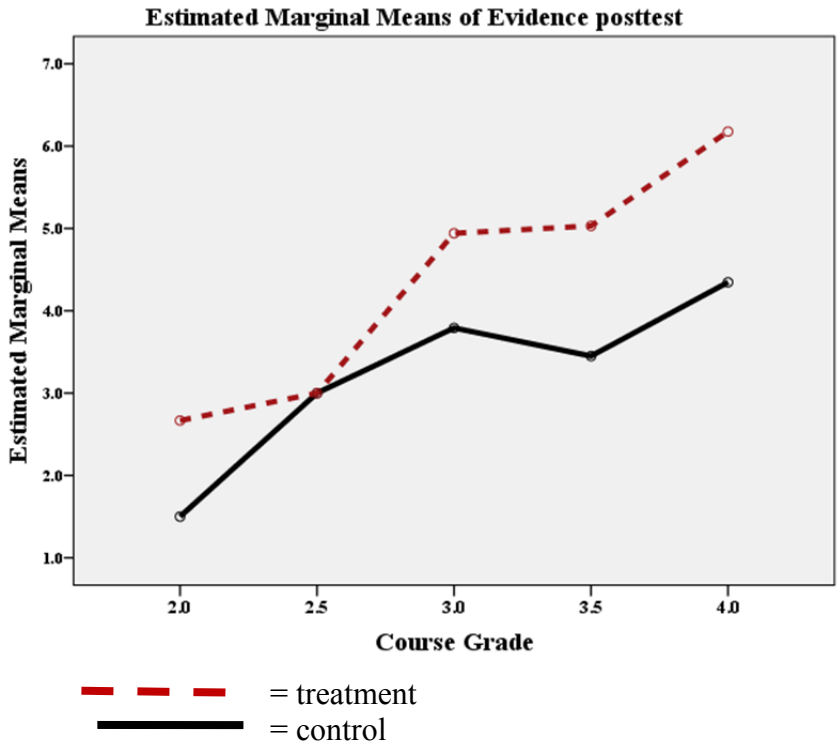


Figure 5-3. Graph of Course Grade and Evidence Posttest Scores

Citation

An analysis of variance showed that the difference between Instructors' grades (course grade) and the judgments of independent scorers as represented by the Citation posttest, was significant, $F(4,159) = 4.611$, $p \leq .003$ (see Table 5-16). This also indicates that the Citation trait is independent of the more general course grade. This difference remained true even if each treatment group was analyzed independently (see Table 5-17). Plotting the Citation scores for these groups separately showed that the trajectory of the line trended up for students in the treatment group. For the control group, the plot remained flatter (see Figure 5-4). Although students in the control group had a range of course grades similar to that of students in the treatment group, their higher grades showed a less direct relationship to their Citation scores.

Table 5-16. Relationship of Course Grade and Citation Posttest Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	133.441[a]	4	33.360	4.611	.002	.107	18.442	.942
Intercept	1293.255	1	1293.255	178.734	.000	.537	178.734	1.000
CourseGrade	133.441	4	33.360	4.611	.002	.107	18.442	.942
Error	1114.289	154	7.236					
Total	6268.750	159						
Corrected Total	1247.730	158						

a. R Squared = .107 (Adjusted R Squared = .084)

b. Computed using alpha = .05

Table 5-17. Course Grade and Citation Posttest Scores by Treatment Group

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	222.216[a]	9	24.691	3.587	.000	.178
Intercept	1178.360	1	1178.360	171.208	.000	.535
CourseGrade	82.632	4	20.658	3.001	.020	.075
TREATMT	3.505	1	3.505	.509	.477	.003
CourseGrade * TREATMT	27.138	4	6.784	.986	.417	.026
Error	1025.514	149	6.883			
Total	6268.750	159				
Corrected Total	1247.730	158				

This indicates that the Evidence trait is independent of the more general course grade. This difference remained true even if each treatment group was analyzed independently Plotting the Evidence scores for these groups separately showed they followed a parallel trajectory (see Figure 5-4).

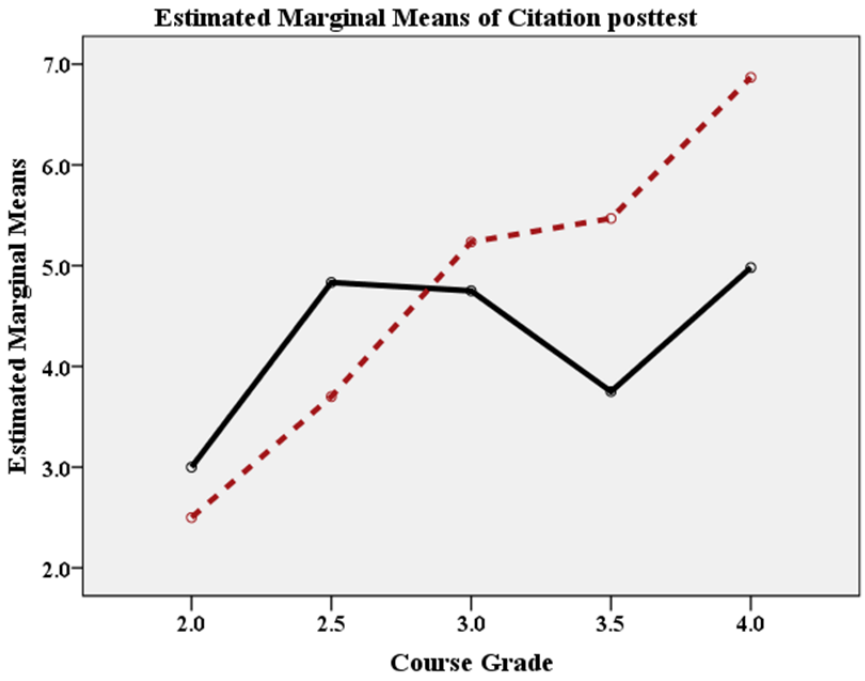


Figure 5-4. Graph of Course Grade and Citation Posttest Scores

Integration

An analysis of variance showed that the difference between Instructors' grades (course grade) and the judgments of independent scorers as represented by the Integration posttest was significant, $F(4,159) = 7.316$, $p \leq .0001$ (See Table 5-18). This indicates that Citation trait is independent of the more general course grade. This difference remained true even if each treatment group was analyzed independently. (see Table 5-19). Plotting the Citation scores for these groups separately showed that for students in the treatment group, the plot line trended up. As course grade increased, Citation posttest score also

increased. For the control group, the plot remained flatter, as illustrated in Figure 5-5). Although students in the control group had a similar range of course grades to students in the treatment group, their higher grades showed a less direct relationship to their Citation scores.

Table 5-18. Relationship of Course Grade and Integration Posttest Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	142.711[a]	4	35.678	7.316	.000	.160	29.265	.996
Intercept	755.012	1	755.012	154.827	.000	.501	154.827	1.000
CourseGrade	142.711	4	35.678	7.316	.000	.160	29.265	.996
Error	750.981	154	4.877					
Total	4176.750	159						
Corrected Total	893.692	158						

a. R Squared = .160 (Adjusted R Squared = .138)
b. Computed using alpha = .05

Table 5-19. Course Grade and Integration Post-test Scores by Treatment Group

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	289.876[a]	9	32.208	7.948	.000	.324	71.531	1.000
Intercept	674.990	1	674.990	166.563	.000	.528	166.563	1.000
CourseGrade	73.690	4	18.423	4.546	.002	.109	18.184	.938
TREATMT	4.012	1	4.012	.990	.321	.007	.990	.167
CourseGrade * TREATMT	40.533	4	10.133	2.501	.045	.063	10.002	.700
Error	603.816	149	4.052					

Total	4176.750	159
Corrected Total	893.692	158

a. R Squared = .324 (Adjusted R Squared = .284)
b. Computed using alpha = .05

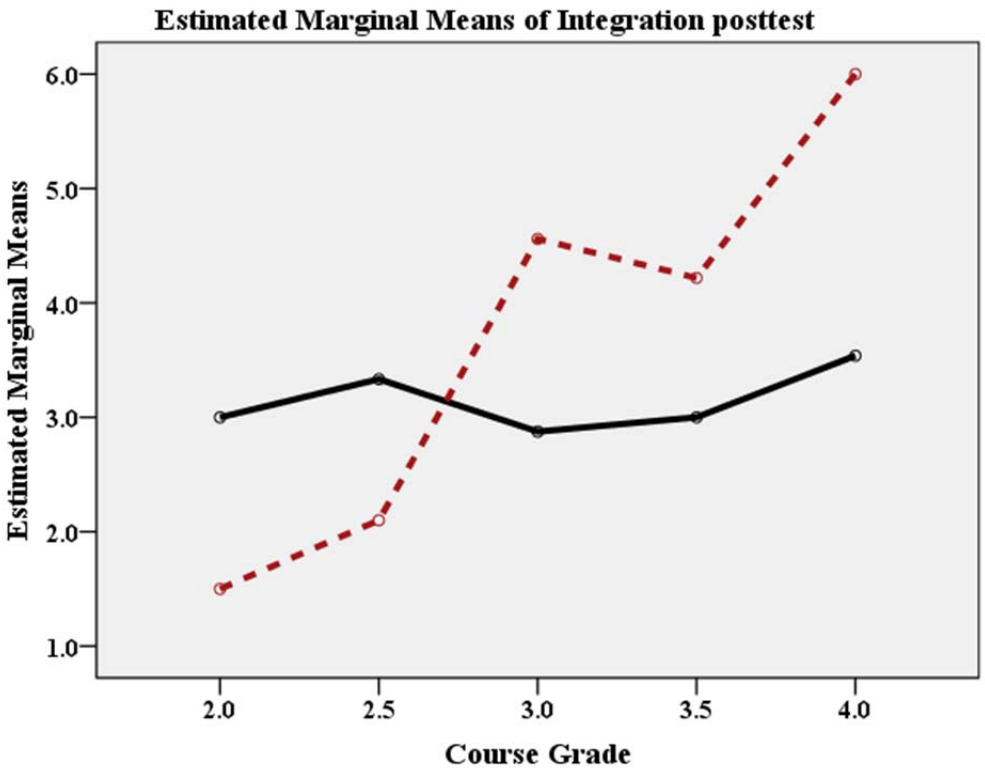


Figure 5-5. Graph of Course Grade and Integration Posttest Scores

Writing

An analysis of variance showed that the difference between Instructors' grades (Course grade) and the judgments of independent scorers as represented by the Writing posttest, was significant, $F(4,159) = 7.975$, $p \leq .0001$ (see Table 5-20). This also indicates that the Writing trait is independent of the more general course grade. This difference remained true even if each treatment group was analyzed independently (see Table 5-21). Plotting the Writing scores for these groups separately showed that the trajectory of the line trended up for students in the treatment and control groups (see Figure 5-6).

Table 5-20. Relationship of Course Grade and Writing Posttest Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	114.002[a]	4	28.501	7.975	.000	.172	31.899	.998
Intercept	875.937	1	875.937	245.098	.000	.614	245.098	1.000
CourseGrade	114.002	4	28.501	7.975	.000	.172	31.899	.998
Error	550.369	154	3.574					
Total	4221.000	159						
Corrected Total	664.371	158						

a. R Squared = .172 (Adjusted R Squared = .150)

b. Computed using alpha = .05

Table 5-21. Course Grade and Writing Posttest Scores by Treatment Group

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	149.287[a]	9	16.587	4.798	.000	.225	43.185	.999
Intercept	819.678	1	819.678	237.111	.000	.614	237.111	1.000
CourseGrade	75.326	4	18.832	5.447	.000	.128	21.790	.972
TREATMT	.003	1	.003	.001	.977	.000	.001	.050
CourseGrade * TREATMT	15.730	4	3.932	1.138	.341	.030	4.550	.351
Error	515.084	149	3.457					
Total	4221.000	159						
Corrected Total	664.371	158						

a. R Squared = .225 (Adjusted R Squared = .178)

b. Computed using alpha = .05

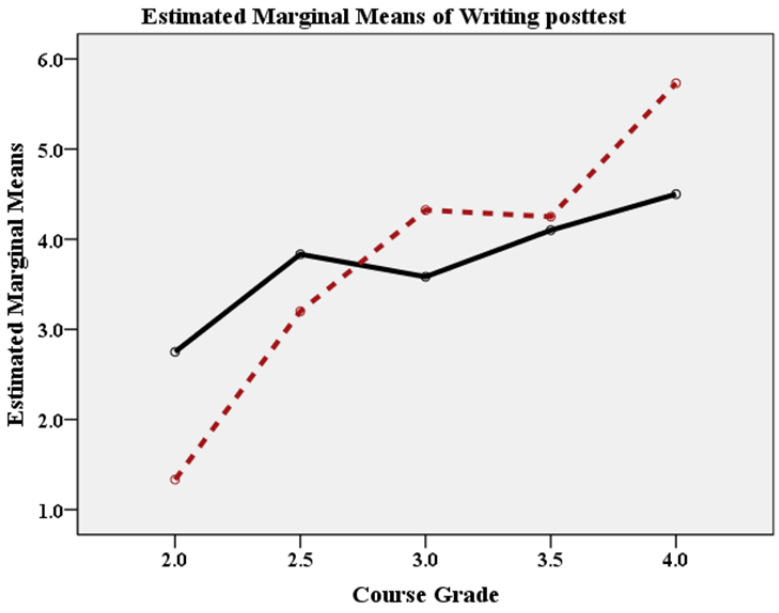


Figure 5-6. Graph of Course Grade and Writing Posttest Scores

Holistic

An analysis of variance showed that the difference between Instructors' grades (Course grade) and the judgments of independent scorers as represented by the Holistic posttest, was significant, $F(4,159) = 8.057$, $p \leq .0001$ (see Table 5-22). This also indicates that the Holistic trait is independent of the more general course grade. This difference remained true even if each treatment group was analyzed independently (see Table 5-23). Plotting the Holistic scores for these groups separately showed that the trajectory of the line trended up for students in the treatment and control groups (see Figure 5-7).

Table 5-22. Relationship of Course Grade and Holistic Posttest Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	151.139[a]	4	37.785	8.057	.000	.173	32.227	.998
Intercept	581.650	1	581.650	124.024	.000	.446	124.024	1.000
CourseGrade	151.139	4	37.785	8.057	.000	.173	32.227	.998
Error	722.231	154	4.690					
Total	3628.373	159						
Corrected Total	873.370	158						

a. R Squared = .173 (Adjusted R Squared = .152)

b. Computed using alpha = .05

Table 5-23. Course Grade and Holistic Posttest Scores by Treatment Group

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	289.218[a]	9	32.135	8.197	.000	.331	73.771	1.000
Intercept	499.922	1	499.922	127.515	.000	.461	127.515	1.000
CourseGrade	87.048	4	21.762	5.551	.000	.130	22.203	.975
TREATMT	13.012	1	13.012	3.319	.070	.022	3.319	.441
CourseGrade * TREATMT	23.950	4	5.987	1.527	.197	.039	6.109	.464
Error	584.152	149	3.920					
Total	3628.373	159						
Corrected Total	873.370	158						

a. R Squared = .331 (Adjusted R Squared = .291)

b. Computed using alpha = .05

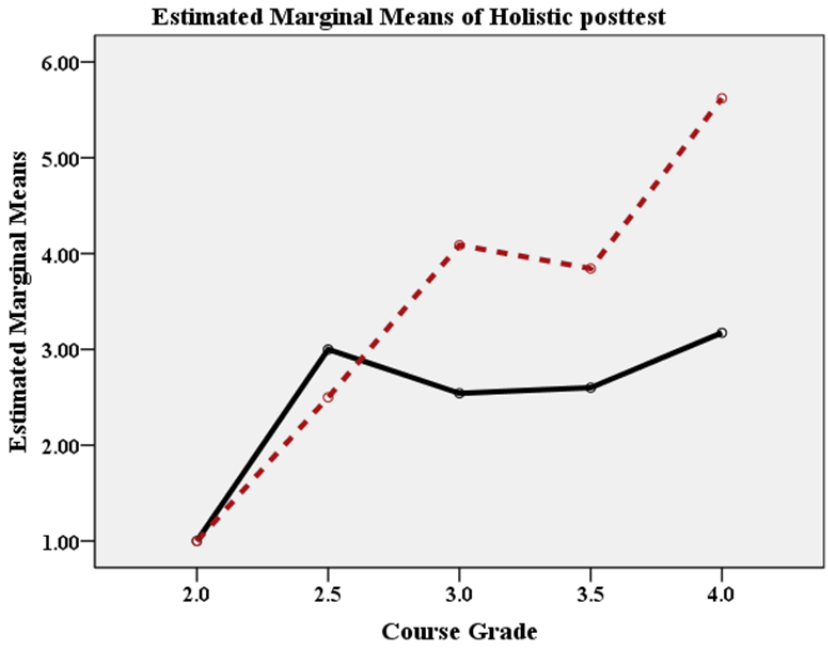


Figure 5-7. Graph of Course Grade and Holistic Posttest Scores

The course grades and information literacy scores might be expected to follow a similar trajectory, however this was not borne out by the data. High scoring students ended the semester with higher course grades than their information literacy scores might indicate. This may be explained by the fact that many of the assignments were graded as low-stakes formative assessments rather than measures of achievement. Making a sincere effort and handing in the work was frequently more than adequate to receive a good grade.

Course Grades and GPA

Validity is also supported by findings that variables that should correlate, do correlate. In Table 5-24, GPA as a measure of overall level of achievement thus far in college correlates significantly with pre- and posttest information literacy scores. The degree of information literacy among students in this study is similar to their cumulative level of academic achievement over the first two or three years of college. This is expected and helps give validity to the diagnostic essay as one more measure of academic achievement (See Table 5-24). Similar correlations were run on SAT scores and information literacy variables, but there was too much missing data among SAT variables, so no analysis could be undertaken.

Table 5-24. GPA Correlations with Information Literacy Variables

		E1 evi	E2 cit	E3 int	E4 writ	E5 hol	T1 Evi	T2 Cit	T3 Int	T4 Writ	T5 Hol
GPA	Pearson	.404**	.264**	.418**	.412**	.423**	.494**	.419**	.502**	.447**	.458**
	Sig.*	.000	.006	.000	.000	.000	.000	.000	.000	.000	.000

*Sig. (2-tailed)

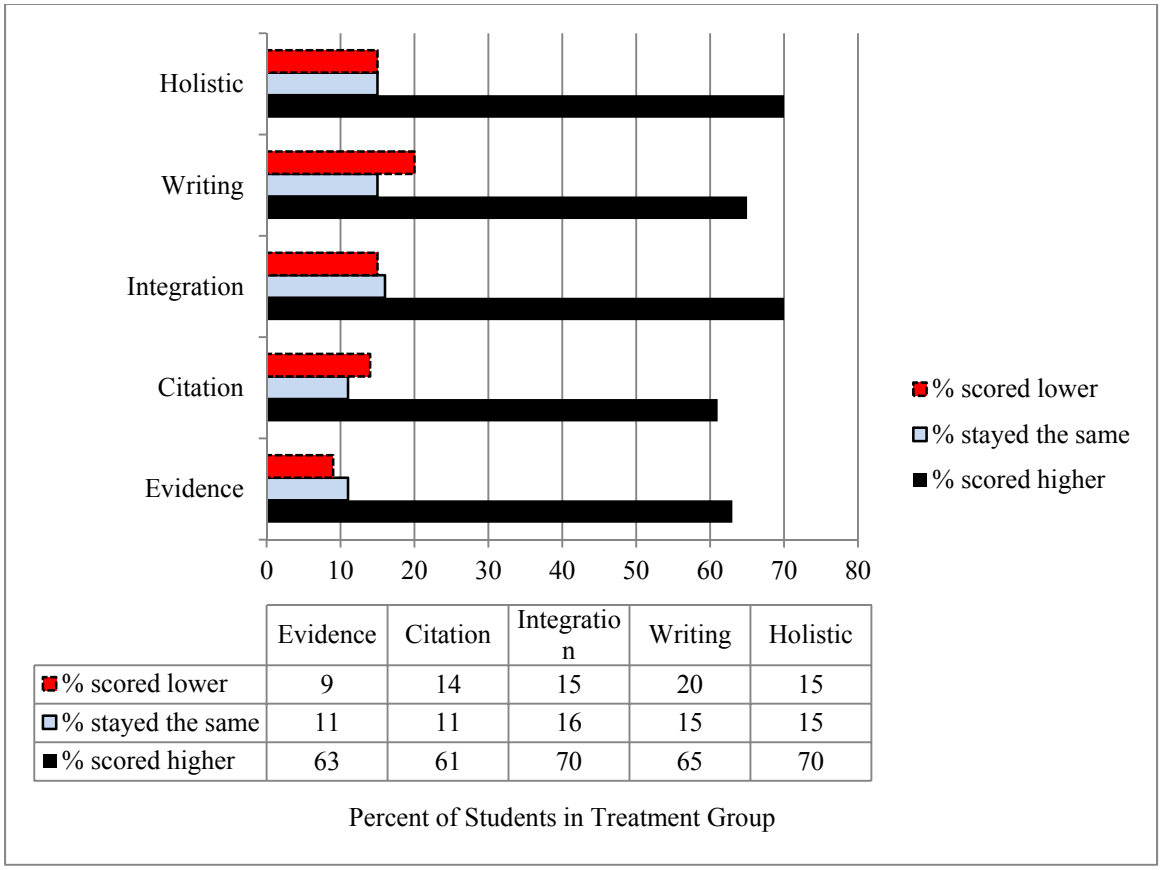
5.5 Findings. Research Question 2. Hypothesis 2.1. Improvement.

Will the intervention designed by the collaborating researcher and instructor improve students' ability to use researched information effectively in their written work?

Hypothesis 2.1: There will be a statistically significant improvement in student performance on the research essay following the intervention specifically targeted at developing the component higher-order information literacy skills.

5.5.1 Comparing Pre- and Posttests Scores

For students in the treatment group, the average score on all five measures of information literacy improved over the course of the semester. In each category, over 60% of the students improved their scores as indicated in Figure 5-8.



	Evidence	Citation	Integration	Writing	Holistic
% scored lower	9	14	15	20	15
% stayed the same	11	11	16	15	15
% scored higher	63	61	70	65	70

Figure 5-8. Means Improved: Pretest Compared to Posttest

The change in distribution of scores from pretest to posttest indicates the general improvement for each pair as in the example of Evidence scores (See Figure 5-9).

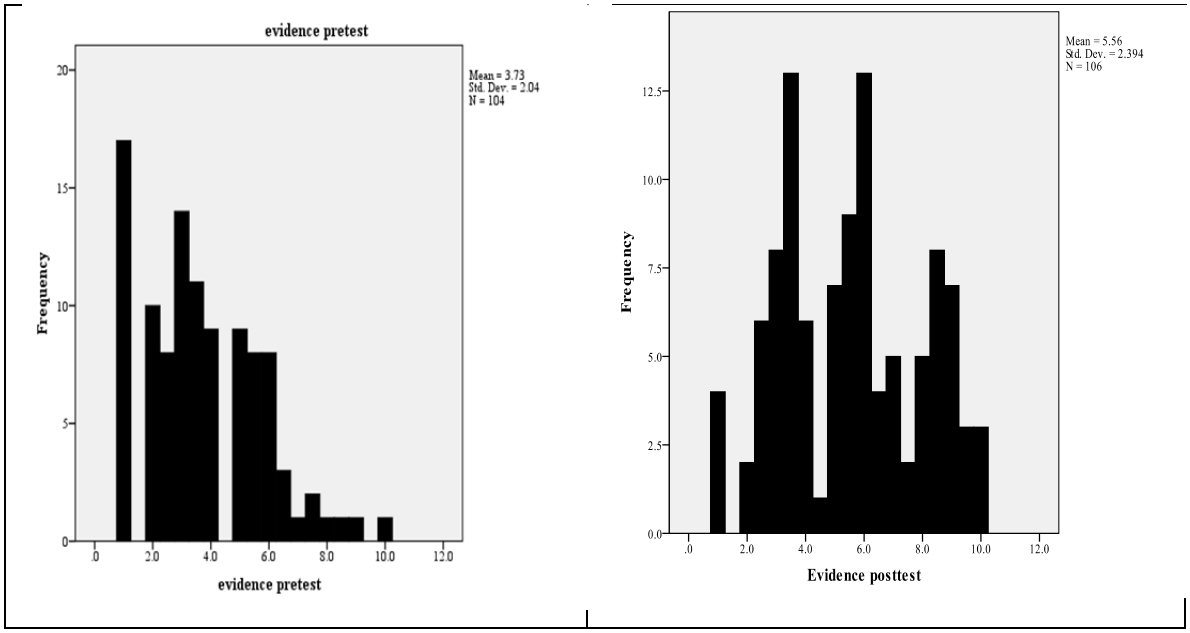


Figure 5-9 Distribution of Evidence Scores Pre and Post

The mean score was highest for Citation following the treatment, followed by Evidence, Integration, Writing, and Holistic. Evidence of Research scores improved from the pretest mean (M = 3.759, SD = 2.0911) to the end of the semester posttest (M = 5.538, SD = 2.3812). Citation improved from the pretest mean (M = 4.759, SD = 2.6378) to the end of the semester posttest (M = 6.151, SD = 2.6342). Integration improved from the pretest mean (M = 3.599, SD = 1.9352) to the posttest (M = 5.208, SD = 2.4357). Writing improved from the pretest mean (M = 4.038, SD = 1.7643) to the posttest (M = 5.057, SD = 2.1551). The Holistic mean improved from the pretest (M = 3.453) to the posttest mean (M = 4.838, SD = 1.7369). (See Table 5-25). Figure 5-10 illustrates the difference. This suggests that the students benefited overall from the intervention.

Specifically, our results suggest that targeting component information literacy skills with specialized instruction and assignments throughout the semester may indeed improve them. As well, a strong relationship exists between the quality of work a student produced upon entering the course with work produced by that student at the end of the term (see Section 5.3). Thus, the diagnostic essay seems to be a good predictor of future performance in information literacy.

Table 5-25. Means Improved: Data on Pretest Compared to Posttest

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Evidence pretest	3.759	106	2.0911	.2031
	Evidence posttest	5.538	106	2.3812	.2313
Pair 2	Citation pretest	4.759	106	2.6378	.2562
	Citation posttest	6.151	106	2.6342	.2559
Pair 3	Integration pretest	3.599	106	1.9352	.1880
	Integration posttest	5.208	106	2.4357	.2366
Pair 4	Writing pretest	4.038	106	1.7643	.1714
	Writing posttest	5.057	106	2.1551	.2093
Pair 5	Holistic pretest	3.453	106	1.7369	.1687
	Holistic posttest	4.838	106	2.3565	.2289

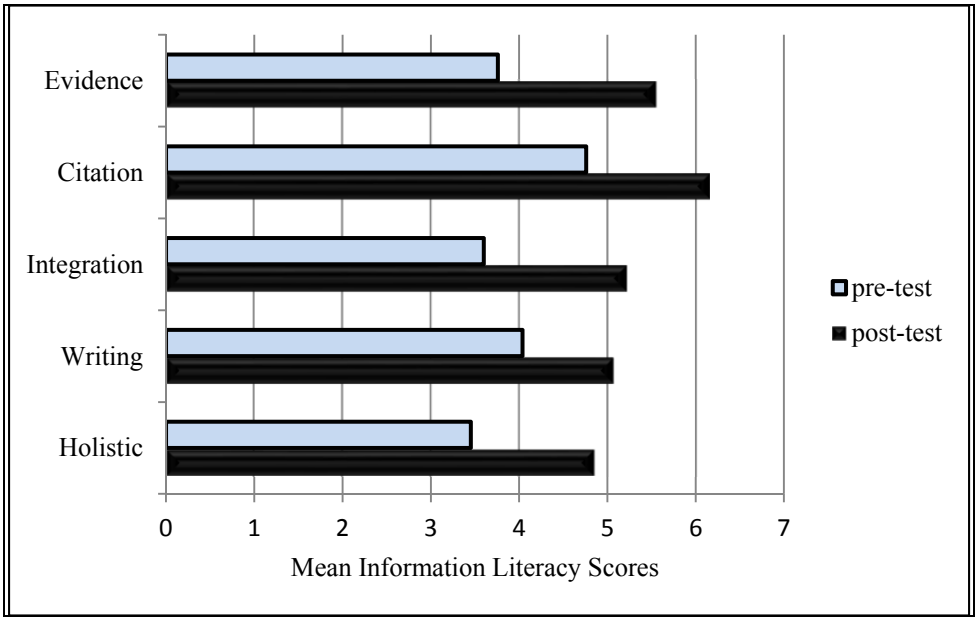


Figure 5-10. Means Improved: Pretest Compared to Posttest

5.5.2 Significance and Effect Size

A repeated measures ANOVA was conducted for each of the information literacy measures to assess whether the differences were significant between the average scores of the pretests and the posttests and to assess the effect sizes. All were computed using a probability level of alpha = .05.

Evidence of Research

The Wilks's Lambda and several other multivariate tests of within subjects effects report the same F's (F = 73.782), and are significant $p \leq .0001$) for within-subject effects for the Evidence variable. This indicates that there is a significant difference in the scores following the instruction, and the effect size (eta squared = .413) is substantive and important for the Evidence criterion (see Table 5-26).

Table 5-26. Test of Within-Subject Effects Pre- to Posttest for Evidence of Research

Effect		Value	F	Hypo-thesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Powerb
Evi-dence	Pillai's Trace	.413	73.782a	1.000	105	.000	.413	73.782	1.000
	Wilks's Lambda	.587	73.782a	1.000	105	.000	.413	73.782	1.000
	Hotellin g's Trace	.703	73.782a	1.000	105	.000	.413	73.782	1.000
	Roy's Largest Root	.703	73.782a	1.000	105	.000	.413	73.782	1.000

a. Exact statistic
b. Computed using alpha = .05
c. Design: Intercept
Within Subjects Design: evidence

Citation

The Wilks's Lambda and several others report the same F's (F = 36.682), and are significant ($p \leq .0001$) for within subject effects for the Citation variable. This indicates that there is a significant difference in the scores following the instruction, and the effect size (eta squared = .259) is meaningful for the Citation criterion (see Table 5-27**).**

Table 5-27. Test of Within-Subject Effects Pre- to Posttest for Citation

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Cita-tion	Pillai's Trace	.259	36.682[a]	1.000	105	.000	.259	36.682	1.000
	Wilks's Lambda	.741	36.682[a]	1.000	105	.000	.259	36.682	1.000
	Hotelling's Trace	.349	36.682[a]	1.000	105	.000	.259	36.682	1.000
	Roy's Largest Root	.349	36.682[a]	1.000	105	.000	.259	36.682	1.000

a. Exact statistic
b. Computed using alpha = .05
c. Design: Intercept
Within Subjects Design: citation

Integration

The Wilks's Lambda and several others report the same F's (F = 61.661), and are significant ($p \leq .0001$) for within subject effects for the Integration variable. This indicates that there is a significant difference in the scores following the instruction, and the effect size (eta squared = .370) is important for the Integration criterion (see Table 5-28).

Table 5-28. Test of Within-Subject Effects Pre- to Posttest for Integration

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Integration	Pillai's Trace	.370	61.661[a]	1.000	105	.000	.370	61.661	1.000
	Wilks's Lambda	.630	61.661[a]	1.000	105	.000	.370	61.661	1.000
	Hotelling's Trace	.587	61.661[a]	1.000	105	.000	.370	61.661	1.000
	Roy's Largest Root	.587	61.661[a]	1.000	105	.000	.370	61.661	1.000

a. Exact statistic
b. Computed using alpha = .05
c. Design: Intercept
Within Subjects Design: integration

Writing

The Wilks's Lambda and several others report the same F's (F = 30.937), and are significant ($p \leq .0001$) for within-subject effects for the Writing variable. This indicates that there is a significant difference in the scores following the instruction, and the effect size (eta squared = .228) is moderate for the Integration criterion (see Table 5-29).

Table 5-29. Test of Within-Subject Effects Pretest to Posttest for Writing

Effect		Value	F	Hypo-thesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Writ-ing	Pillai's Trace	.228	30.937[a]	1.000	105	.000	.228	30.937	1.000
	Wilks's Lambda	.772	30.937[a]	1.000	105	.000	.228	30.937	1.000
	Hotelling's Trace	.295	30.937[a]	1.000	105	.000	.228	30.937	1.000
	Roy's Largest Root	.295	30.937[a]	1.000	105	.000	.228	30.937	1.000

a. Exact statistic
b. Computed using alpha = .05
c. Design: Intercept
Within Subjects Design: writing

Holistic

The Wilks's Lambda and several others report the same F's (F = 53.709), and are significant ($p \leq .0001$) for within subject effects for the Holistic variable. This indicates that there is a significant difference in the scores following the instruction, and the effect size (eta squared = .338) is important for the Holistic criterion (see Table 5-30).

Table 5-30. Test of Within-Subject Effects Pretest to Posttest for Holistic

Effect		Value	F	Hypo-thesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Hol-istic	Pillai's Trace	.338	53.709[a]	1.000	105	.000	.338	53.709	1.000
	Wilks's Lambda	.662	53.709[a]	1.000	105	.000	.338	53.709	1.000
	Hotelling's Trace	.512	53.709[a]	1.000	105	.000	.338	53.709	1.000
	Roy's Largest Root	.512	53.709[a]	1.000	105	.000	.338	53.709	1.000

a. Exact statistic
b. Computed using alpha = .05
c. Design: Intercept
Within Subjects Design: holistic

The Repeated Measures ANOVA for each dependent variable indicates that on average, students in the treatment group improved following the instruction and that there were meaningful effect sizes for each measure of information literacy as reported in Table 5-31. These results confirm expectations, and suggest that the information literacy course work had a significant impact on the quality of students' researched writing by the end of

the term. The students had measurable and significant improvement in each of the information literacy skills.

Table 5-31. Consolidated Results of Repeated Measures ANOVA for all Information Literacy Criteria

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Evidence	Pillai's Trace	.413	73.782[a]	1.000	105	.000	.413	73.782	1.000
Citation	Pillai's Trace	.259	36.682[a]	1.000	105	.000	.259	36.682	1.000
Integration	Pillai's Trace	.370	61.661[a]	1.000	105	.000	.370	61.661	1.000
Writing	Pillai's Trace	.228	30.937[a]	1.000	105	.000	.228	30.937	1.000
Holistic	Pillai's Trace	.338	53.709[a]	1.000	105	.000	.338	53.709	1.000

a. Exact statistic
b. Computed using alpha = .05
c. Design: Intercept, Within Subjects Design

Effect Size

Effect size is also important to fully understanding the results. A result may be significant, but the effect size may be too small to support the importance and impact of the findings. Large effect sizes (partial eta squared) for Evidence, Integration, and Holistic are shown in Table 5-31. Glass's Δ was also calculated to describe the effect size independent of the units of measurement. This is useful in understanding the results in comparison to effect sizes for other types of variables that affect educational outcomes. Using 5.5 as the hypothetical mean, the effect sizes for each of the information literacy variables following the treatment were calculated as reported in Table 5-32.

Effect sizes for Citation and Holistic are substantially lower than the others. The effect on Citation may be lower because students came in with a good degree of knowledge about citing sources and thus had less room for improvement. The Holistic trait measured the student's level of sophistication in academic researched writing, a complex skill combination perhaps requiring more than one semester to make substantial improvement. Several of these measures (Evidence, Integration, and Writing) approach 1.0. These may be categorized as large effect sizes according to Cohen (1992). However, it has been suggested that rather than accept these somewhat arbitrary categories, that effect sizes be compared with average effect sizes for similar research (Hattie, 1999, 2005; Walberg, 1984). This is not only useful as a way of understanding the results, but also may aid practitioners in focusing on factors that produce the biggest gains.

Table 5-32. Effect Sizes for Treated Group

Score	Glass's Δ
Evidence	.832
Citation	.281
Integration	.982
Writing	.829
Holistic	.381

Hattie has analyzed thousands of studies of learning and suggested that .4 is the average effect size of educational variables that affect learning. For example, for comparison, he found the effect of advancing learning by one grade level is 1.0, whereas the effect of student maturation is 0.1. In light of these numbers, our finding of effect sizes of from .281 to .982 is notable given that this represents the learning over one semester of 15 weeks. If such gains could be achieved over several semesters while in college, it would result in a great improvement overall in the research skills of these

students by graduation. These measures were calculated by Hattie for K–12 students. However, research on undergraduates has shown even smaller effect sizes for college impact variables. In their meta-analysis of college impact research on undergraduates Pascarella and Terenzini (2005) calculated the average improvement to be .50 standard deviation from freshman to senior year. Arum and Roksa (2011) showed a similar gain in critical thinking, as measured by the Collegiate Learning Assessment over four years of college. In addition, few information literacy studies report effect sizes, and none could be found that used constructed-response tests. Koufogiannakis (2006) lacked enough data to compare effect-sizes in her meta-analysis of effective methods of teaching information literacy. Burkhardt (2007) claimed to have found a high effect size of a three-credit information literacy course in a pre/post measurement using a multiple choice test, but did not report the specific evidence to support that finding. Thus, these findings are important benchmarks of the size of the effect of a library skills intervention on ACRL information literacy standards of performance.

5.5.3 Gain Scores

Another way to look at the degree of improvement during the semester is to calculate the differences between the individual student's scores on the pretest and the posttest as variables (GAIN scores). The mean difference was positive for each trait (Evidence M = 1.778, SD = 2.131; Citation M = 1.3915, SD = 2.365; Integration M = 1.608, SD = 2.109; Writing M = 1.019, SD = 4.448; Holistic M = 1.385, SD = 3.788 as reported in Table 5-33. These scores were found to correlate well with all the diagnostic traits except the holistic score. There was a positive correlation between the pretest for Evidence of Research (M = 3.759, SD = 2.0911) and the corresponding GAIN Score (M

= 1.778, SD = 2.365), r = .364, $p \leq .001$, n = 106. There was a positive correlation between the pretest for Citation (M = 4.759, SD = 2.638) and the corresponding GAIN Score (M = 1.392, SD = 2.365), r = .450, $p \leq .001$, n = 106. There was a positive correlation between the pretest for Integration (M=3.599, SD=1.935) and the corresponding GAIN Score (M = 1.609, SD = 2.109), r = .277, $p \leq .001$, n = 106. There was a positive correlation between the pretest for Writing (M = 4.038, SD = 1.764) and the corresponding GAIN Score (M = 1.019, SD = 1.886), r = .304, $p \leq .001$, n = 106. The correlation coefficient was not significant between the pretest for Holistic (M = 3.453, SD = 1.737) and the corresponding GAIN Score (M = 1.385, SD = 1.946), r = .185, $p \leq .001$, n = 106. (See Table 5-34).

Table 5-33. Compare Means for Pretest and Gain Scores

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Evidence pre-test	3.759	106	2.091	0.203
	GAIN on Evidence	1.778	106	2.131	0.207
Pair 2	Citation pre-test	4.759	106	2.638	0.256
	GAIN on Citation	1.392	106	2.365	0.230
Pair 3	Integration pre-test	3.599	106	1.935	0.188
	GAIN on Integration	1.609	106	2.109	0.205
Pair 4	Writing pre-test	4.038	106	1.764	0.171
	GAIN on Writing	1.019	106	1.886	0.183
Pair 5	Holistic pre-test	3.453	106	1.737	0.169
	GAIN on holistic	1.385	106	1.946	0.189

Table 5-34. Correlations between Diagnostic Test Scores and Gain Scores

		GAIN Scores for				
		Evidence	Citation	Integration	Writing	Holistic
Evidence pretest	Pearson Correlation	-.364**	-.138	-.104	-.048	-.006
	Sig. (2-tailed)	.000	.158	.290	.623	.951
	Sum of Squares and Cross-products	-170.403	-71.767	-47.983	-20.019	-2.573
	Covariance	-1.623	-.683	-.457	-.191	-.025
Citation pretest	Pearson Correlation	-.038	-.450**	-.095	.032	-.038
	Sig. (2-tailed)	.699	.000	.335	.748	.698
	Sum of Squares and Cross-products	-22.403	-294.767	-55.233	16.481	-20.523
	Covariance	-.213	-2.807	-.526	.157	-.195
Integration pretest	Pearson Correlation	-.123	-.167	-.277**	-.045	-.091
	Sig. (2-tailed)	.208	.088	.004	.648	.354
	Sum of Squares and Cross-products	-53.422	-80.111	-118.639	-17.198	-35.921
	Covariance	-.509	-.763	-1.130	-.164	-.342
Writing pretest	Pearson Correlation	-.167	-.203*	-.175	-.304**	-.139
	Sig. (2-tailed)	.086	.037	.074	.002	.155
	Sum of Squares and Cross-products	-66.113	-89.066	-68.184	-106.325	-50.167
	Covariance	-.630	-.848	-.649	-1.013	-.478
Holistic pretest	Pearson Correlation	-.162	-.157	-.189	-.140	-.185
	Sig. (2-tailed)	.096	.108	.052	.152	.057
	Sum of Squares and Cross-products	-63.108	-67.792	-72.708	-48.156	-65.698
	Covariance	-.601	-.646	-.692	-.459	-.626

In summary, the percentage of students in the treatment group who improved following the information literacy instruction was high. As well, the degree of improvement overall was great. Because this transpired over the course of only 15 weeks, it is unlikely that maturational or other factors played a significant role. Therefore, it is likely that something about the instruction and assignments made a difference.

5.6 Findings. Research Question 2. Hypothesis 2.2. Treatment versus Control

Hypothesis 2.2. Students in the treatment group will perform significantly better than those in the control group.

In order to determine whether or not the students in the treatment group significantly outperformed the students in the control group, an Independent Samples *t*-test was conducted. The results indicated that the mean for each information literacy trait on the posttest was higher for the 108 students in the treatment group than the means for the 53 students in the control group as reported in Table 5-35. The mean difference between the groups ranged from 0.9 to 2.0 for the five posttest criteria. The scores improved on average about 10% in each category.

Table 5-35. Means for Treatment and Control Groups

Essay Scores	Treatment Group	N	Mean	Std. Deviation	Std. Error Mean
Evidence	treatment	108	5.565	2.376	.229
post-test	control	53	3.868	2.031	.279
Citation	treatment	108	6.153	2.610	.251
post-test	control	53	4.613	2.941	.404
Integration	treatment	108	5.181	2.421	.233
post-test	control	53	3.255	1.669	.229
Writing	treatment	108	5.032	2.142	.206
post-test	control	53	4.113	1.695	.233
Holistic	treatment	108	4.818	2.339	.225
post-test	control	53	2.830	1.701	.234

There was a significant difference between the scores of the treated students for Evidence of Research (M = 5.565, SD = 2.376) and the control group (M = 3.868, SD = 2.031); $t(159) = 4.459$, $p \leq .0001$. There was a significant difference for the Citation variable between the treated students (M = 6.153, SD = 2.610) and the control group (M = 4.613, SD = 2.941; $t(159) = 3.372$), $p \leq .002$. There was a significant difference for the Integration variable between the treated students (M = 5.181, SD = 2.421) and those in the control group (M = 3.255, SD = 1.669); $t(159) = 5.892$, $p \leq .0001$. There was a significant difference for the Writing variable between the treated students (M = 5.032, SD = 2.142) and those in the control group (M = 4.113, SD = 1.694); $t(159) = 2.596$, $p \leq .005$. There was a significant difference for the Holistic variable between the treated students (M = 4.818, SD = 2.339) and the control group (M = 2.830, SD = 1.701); $t(159) = 6.127$, $p \leq .0001$. The assumption of homogeneity of variance was assessed by the Levene test. The tests for Evidence of Research ($F = 2.162$, $p = .143$) and Citation ($F = 2.763$, $p = .098$) indicated no significant violation of the equal variance assumption, therefore the *equal variances assumed* version of the test was used. Levene's test for Integration ($F = 18.337$, $p = .000$), Writing ($F = 8.120$, $p = .005$) and Holistic ($F = 11.919$, $p = .001$) were significant, therefore the *equal variances not assumed* version of the test was used for those variables (see Table 5-36). These results suggest that by following the information literacy instructional program devised by the researcher and the collaborating instructor, students achieved higher degrees of information literacy than students who did not receive this training. The positive effect size, as indexed by η^2, was large for Evidence of Research (.33), medium for Integration (.19) and Holistic (.22), and small for

Citation (.07) and Writing (.06). The sample size was more than adequate to provide good statistical power.

It is interesting to note that the means for the holistic posttest score for the treatment group (M = 4.818) and the control group (M = 2.830) are both the lowest of any trait. This provides evidence of the validity of the holistic score, because it should be the most difficult to achieve. It is possible that a student may perform adequately on the other individual traits and still display weakness overall.

Table 5-36. Results of the Independent Samples t-Test Comparing Treatment and Control Groups

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
T1eviadjtotal	Equal variances assumed	2.162	.143	4.459	159	.000	1.6969	.3805	.9453	2.4485
	Equal variances not assumed			4.704	119.159	.000	1.6969	.3607	.9826	2.4112
T2citadjtotal	Equal variances assumed	2.763	.098	3.372	159	.001	1.5396	.4566	.6378	2.4414
	Equal variances not assumed			3.237	93.184	.002	1.5396	.4757	.5950	2.4842
T3intadjtotal	Equal variances assumed	18.337	.000	5.211	159	.000	1.9258	.3696	1.1960	2.6557
	Equal variances not assumed			5.892	141.526	.000	1.9258	.3268	1.2797	2.5719
T4writadjtotal	Equal variances assumed	8.120	.005	2.731	159	.007	.9192	.3366	.2544	1.5840
	Equal variances not assumed			2.956	127.462	.004	.9192	.3109	.3039	1.5345
T5holadjtotal	Equal variances assumed	11.919	.001	5.509	159	.000	1.98787	.36084	1.27522	2.70052
	Equal variances not assumed			6.127	136.268	.000	1.98787	.32445	1.34626	2.62948

A one-way analysis of variance (ANOVA) was also calculated to obtain the F ratio as another measure of the differences between the treatment and control groups. The ANOVA identifies the part of each individual score that is associated with membership in the two groups (treatment or control) and the part that is not, and attributable to "error." There was a high level of significant difference between the groups $F(1,160) = 19.884$, $p \leq .0001$ for Evidence of Research, $F = 11.369$, $p = .001$ for Citation, $F = 27.156$, $p = .001$ for Integration, $F = 7.458$, $p = .007$ for Writing, and $F = 30.350$, $p = .007$ for the Holistic variable (see Table 5-37). Again, the results show that on average students in the treatment group improved significantly over the control group. Effect sizes as measured by adjusted r squared: Evidence of Research ($r^2 = 111$), Citation ($r^2 = .067$), Integration ($r^2 = .146$), Writing ($r^2 = .045$), and Holistics ($r^2 = .160$).

Table 5-37. Comparison of Treatment versus Control ANOVA with Effect Sizes

Source	Posttest Essay Scores (DVs)	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	Evidence	102.372[a]	1	102.372	19.884	.000	.111	19.884	.993
	Citation	84.270[c]	1	84.270	11.369	.001	.067	11.369	.918
	Integration	130.595[d]	1	130.595	26.574	.000	.143	26.574	.999
	Writing	30.040[e]	1	30.040	7.458	.007	.045	7.458	.775
	Holistic	140.491[f]	1	140.491	30.350	.000	.160	30.350	1.000
Intercept	Evidence	3163.366	1	3163.366	614.417	.000	.794	614.417	1.000
	Citation	4120.798	1	4120.798	555.943	.000	.778	555.943	1.000
	Integration	2524.167	1	2524.167	513.625	.000	.764	513.625	1.000
	Writing	2973.717	1	2973.717	738.255	.000	.823	738.255	1.000
	Holistic	2079.683	1	2079.683	449.268	.000	.739	449.268	1.000
Treatment	Evidence	102.372	1	102.372	19.884	.000	.111	19.884	.993
	Citation	84.270	1	84.270	11.369	.001	.067	11.369	.918
	Integration	130.595	1	130.595	26.574	.000	.143	26.574	.999
	Writing	30.040	1	30.040	7.458	.007	.045	7.458	.775
	Holistic	140.491	1	140.491	30.350	.000	.160	30.350	1.000
Error	Evidence	818.622	159	5.149					
	Citation	1178.550	159	7.412					
	Integration	781.392	159	4.914					
	Writing	640.457	159	4.028					
	Holistic	736.019	159	4.629					
Total	Evidence	4956.000	161						
	Citation	6395.000	161						
	Integration	4231.000	161						
	Writing	4272.250	161						
	Holistic	3667.623	161						
Corrected Total	Evidence	920.994	160						
	Citation	1262.820	160						
	Integration	911.988	160						
	Writing	670.497	160						
	Holistic	876.510	160						

a. R Squared = .111 (Adjusted R Squared = .106)

b. Computed using alpha = .05

c. R Squared = .067 (Adjusted R Squared = .061)

d. R Squared = .143 (Adjusted R Squared = .138)

e. R Squared = .045 (Adjusted R Squared = .039)

f. R Squared = .160 (Adjusted R Squared = .155)

In summary, the level of achievement of students in the treatment group was significantly greater in each category at the end of the semester than that of students in the control group. This lends further support to the conclusion that it was most likely the librarian's information literacy intervention that made a significant difference.

5.6.1 Low Skills/High Skills: Search and Cite versus Integrate and Use

In order to further explore student performance in more detail, an index of lower level information literacy skills was created by averaging Evidence of Research and Citation, and an index of higher level information literacy skills by averaging Integration and Writing scores. The difference in achievement between the treatment and control groups was significant for low and high skills, although the means were higher for students in the treatment group. Table 5-38 highlights the finding that, on average, all the students performed better on the lower level skills than on the higher ones. This further validates the relationship of the criteria to each other. It makes sense that when grouped by difficulty, the search and cite skills have higher scores in the aggregate because they are thought to be "easier" than the integrate and use skills. It is possible that since it takes more time and practice to learn a more complex higher-level skill, these results might be different if the interval between pre- and posttest were longer. This hypothesis can be tested in the future.

Table 5-38. Means for Low Skills versus High Skills

	Treatment Group	N	Mean	Std. Deviation	Std. Error Mean
LowAvg	treatment	108	5.8588	2.28191	.21958
posttest	control	53	4.2406	2.05452	.28221
HighAvg	treatment	108	5.1065	2.23168	.21474
posttest	control	53	3.6840	1.53575	.21095

5.7 Findings. Research Question 3. Hypothesis 3.1. Pretest Effect

Hypothesis 3.1. The pretest accounts for a portion of the variance in scores between the pre- and posttest.

Simply taking the pretest itself may affect performance on the posttest, and therefore represents a threat to the validity of the experiment. For this reason, an analysis of covariance (ANCOVA) was conducted to remove the effect of taking the pretest from the posttest performance. Only two pretest traits had a significant influence on the corresponding posttest traits. The Evidence of Research pretest had a small but significant effect on the Evidence of Research posttest ($\eta^2 = .091$, $p = .002$). The Citation pretest trait had a small but significant effect on the Citation posttest ($\eta^2 = .173$, $p \leq .0001$). Because r squared is high overall for all, more than 30% of the variability is explained by the treatment after removing any influence of the pretests measures. (Evidence of Research [$r^2 = .342$, $p \leq .0001$]; Citation [$r^2 = .393$, $p \leq .0001$]; Integration [$r^2 = .333$, $p \leq .0001$]; Writing [$r^2 = .404$), $p \leq .0001$]; Holistic [$r^2 = .369$, $p \leq .0001$]). (see Table 5-39). There are other threats to validity, due to the time that lapsed between the pretest and the posttest. Students may have learned or had experiences outside of this class that affected their knowledge about using external sources in their writing. Or they may approach the assignment with increased maturity simply because they have grown older.

Table 5-39. Variability Explained

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	Evidence posttest	203.677[a]	5	40.735	10.400	.000	.342	52.002	1.000
	Citation posttest	286.044[c]	5	57.209	12.927	.000	.393	64.637	1.000
	Integration posttest	207.562[d]	5	41.512	9.994	.000	.333	49.970	1.000
	Writing posttest	196.814[e]	5	39.363	13.534	.000	.404	67.669	1.000
	Holistic posttest	215.225[f]	5	43.045	11.701	.000	.369	58.505	1.000
Intercept	Evidence posttest	115.169	1	115.169	29.404	.000	.227	29.404	1.000
	Citation posttest	161.226	1	161.226	36.432	.000	.267	36.432	1.000
	Integration posttest	80.004	1	80.004	19.261	.000	.162	19.261	.991
	Writing posttest	68.353	1	68.353	23.502	.000	.190	23.502	.998
	Holistic posttest	50.299	1	50.299	13.673	.000	.120	13.673	.956
Evidence posttest	Evidence posttest	39.025	1	39.025	9.964	.002	.091	9.964	.878
	Citation posttest	.164	1	.164	.037	.848	.000	.037	.054
	Integration posttest	5.787	1	5.787	1.393	.241	.014	1.393	.215
	Writing posttest	2.540	1	2.540	.873	.352	.009	.873	.152
	Holistic posttest	6.005	1	6.005	1.632	.204	.016	1.632	.244
Citation posttest	Evidence posttest	3.386	1	3.386	.864	.355	.009	.864	.151
	Citation posttest	92.286	1	92.286	20.854	.000	.173	20.854	.995
	Integration posttest	3.927	1	3.927	.945	.333	.009	.945	.161
	Writing posttest	3.955	1	3.955	1.360	.246	.013	1.360	.211
	Holistic posttest	1.001	1	1.001	.272	.603	.003	.272	.081
Integration posttest	Evidence posttest	7.740	1	7.740	1.976	.163	.019	1.976	.286
	Citation posttest	3.408	1	3.408	.770	.382	.008	.770	.140
	Integration posttest	8.777	1	8.777	2.113	.149	.021	2.113	.302
	Writing posttest	7.042	1	7.042	2.421	.123	.024	2.421	.338
	Holistic posttest	2.721	1	2.721	.740	.392	.007	.740	.136
Writing posttest	Evidence posttest	.036	1	.036	.009	.924	.000	.009	.051
	Citation posttest	10.096	1	10.096	2.281	.134	.022	2.281	.322
	Integration posttest	.061	1	.061	.015	.904	.000	.015	.052
	Writing posttest	.519	1	.519	.178	.674	.002	.178	.070

	Holistic posttest	.034	1	.034	.009	.924	.000	.009	.051
Holistic	Evidence posttest	.875	1	.875	.223	.637	.002	.223	.075
posttest	Citation posttest	6.589	1	6.589	1.489	.225	.015	1.489	.227
	Integration posttest	2.093	1	2.093	.504	.479	.005	.504	.108
	Writing posttest	1.569	1	1.569	.539	.464	.005	.539	.112
	Holistic posttest	7.427	1	7.427	2.019	.158	.020	2.019	.291
Treatment	Evidence posttest	.000	0	.	.	.	.000	.000	.
	Citation posttest	.000	0	.	.	.	.000	.000	.
	Integration posttest	.000	0	.	.	.	.000	.000	.
	Writing posttest	.000	0	.	.	.	.000	.000	.
	Holistic posttest	.000	0	.	.	.	.000	.000	.
Error	Evidence posttest	391.672	100	3.917					
	Citation posttest	442.541	100	4.425					
	Integration posttest	415.372	100	4.154					
	Writing posttest	290.846	100	2.908					
	Holistic posttest	367.873	100	3.679					
Total	Evidence posttest	3846.000	106						
	Citation posttest	4739.000	106						
	Integration posttest	3497.500	106						
	Writing posttest	3198.000	106						
	Holistic posttest	3064.373	106						
Corrected	Evidence posttest	595.349	105						
Total	Citation posttest	728.585	105						
	Integration posttest	622.934	105						
	Writing posttest	487.660	105						
	Holistic posttest	583.098	105						

a. R Squared = .342 (Adjusted R Squared = .309)
b. Computed using alpha = .05
c. R Squared = .393 (Adjusted R Squared = .362)
d. R Squared = .333 (Adjusted R Squared = .300)
e. R Squared = .404 (Adjusted R Squared = .374)
f. R Squared = .369 (Adjusted R Squared = .338)

Taking the pretest had a small but significant positive effect on student performance on two of the criteria. This is logical because taking a test more than once often results in improved scores, as the success of the SAT preparation industry so well illustrates. However, after removing the effects of the pretest, the difference in treatment still explained a significant amount of the change in scores. Because the pretest seems to function as an "assignment" as well as a test, perhaps it can also be considered as part of the instruction in future research. This result provides important evidence that the use of a pretest should be considered not only as an assessment tool, but also as a teaching technique.

5.8 Findings. Research Question 3. Hypothesis 3.2. Prior Conditions

Hypothesis 3.2. Socioeconomic status and other prior conditions will show a significant correlation with information literacy scores.

Socioeconomic Status

Parental education has been one of the more widely used and stable measures of SES in studies of academic achievement over decades of research (Sirin, 2005; White, 1982). Both Sirin and White's meta-analyses of the relationship of SES and academic achievement found significant correlations. White's results (1918–1975) showed an average correlation of r = .343, whereas Sirin (1990–2000) found it to be r = .299. For this study, SES of the students was measured by whether or not at least one parent graduated from college or higher. Because the accuracy of the socioeconomic data is important, they were obtained from the following three sources and cross-checked: 1) student self-report via survey administered by the author; 2) student self-report on admissions application; and 3) student self-report on a financial aid application. The data

were incomplete in any one source and there were occasional discrepancies among the sources. The questions were asked in the following ways.

1) In the author's survey, the question was stated as follows:

   What is the highest educational level attained by either of your parents (for example, your mother OR your father) or the most significant caretaker who influenced you as you were growing up?

   - 8th grade
   - high school
   - some college
   - graduated college
   - master's level degree completed
   - PhD., MD, or other doctoral level degree completed

2) On the admissions application, the question was stated as follows:

   Did your parents graduate from a four-year college/university?

   Mother □ Yes □ No Father □ Yes □ No

3) On the financial aid application, the question was stated as follows:

   Some states and colleges offer aid based on the level of schooling your parents completed.

   Highest school your father completed:

   □ Middle school/Jr. high □ High school □ College or beyond □ Other/unknown

   Highest school your mother completed:

   □ Middle school/Jr. high □ High school □ College or beyond □ Other/unknown

Of the 62 studies reviewed by Sirin, the most reliable data was reported by parents. Based on Sirin's analysis of the greater effect sizes found when parental reporting was used, an assumption was made that the most accurate reporting was likely to be made on the financial aid application, then on the admissions application, and the least reliable was assumed to be the student survey response. When missing data or conflicts arose, the data were considered in that order. The data were transformed twice: first into a dichotomous variable for each parent to designate whether or not that parent had graduated from College, and then into a dichotomous indicator dividing students into those who were the first in their families to attend college, and those who had at least one parent who had completed a college education or higher.

Students who were the first in their families to go to college made up 40% (n = 64) of the sample, whereas 60% (n = 95) had at least one parent who graduated from college or higher. This is approximately the same distribution as reported by NJIT Institutional Research in 2011 (38% were first in family to attend college, 63% had at least one parent who graduated from college). The distribution of students differed slightly in the treatment and control groups. There is a smaller percentage (34%) of FIF students in the control group than in the treatment group (43%) as reported in Table 5-40.

Table 5-40. Distribution of Students According to Socioeconomic Status

			treatment group		
			control	treatment	Total
SES	FIF	Count	18	46	64
		% within SES	28%	72%	100%
		% within treatment group	34%	43%	40%
		% of Total	11%	29%	40%
	COLL	Count	35	60	95
		% within SES	37%	63%	100%
		% within treatment group	66%	57%	60%
		% of Total	22%	38%	60%
Total		Count	53	106	159
		% within SES	33%	67%	100%
		% within treatment group	100%	100%	100%
		% of Total	33%	67%	100%

An independent samples *t*-test was conducted for first in family students (FIF) as compared to students who had at least one parent who graduated college (COLL). There was no significant difference in average GPA between the two groups (see Table 5-41).

This suggests that by the third year some of the presumed differences may have already been smoothed out. One reason may be that the students in this study were already successful students. The majority of participants were science or engineering majors who were either juniors or seniors. Thus the problem of restricted range may account for these results. Most of the undergraduates who successfully complete the first two years in these fairly demanding majors have already mastered the learning skills need to succeed in college. This problem can be addressed in future studies with a larger sample representing undergraduates to include freshmen through seniors. For further comparison, the researcher attempted to collect accurate SAT scores on Math, Verbal,

and Writing, because these represent skill levels at the time of admission. However, too many of the students' SAT scores were missing to carry out that analysis.

Table 5-41. Difference in GPA between FIF and COLL Students for All Participants

	SES	N	Mean	Std. Deviation	Std. Error Mean
GPA	FIF	63	305.44	51.648	6.507
	COLL	91	303.46	58.373	6.119

		Levene's Test for Equality of Variances		*t*-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Diff-erence	Std. Error Diff-erence	Lower	Upper
GPA	Equal variances assumed	2.266	.134	.217	152	.828	1.983	9.134	-16.062	20.028
	Equal variances not assumed			.222	143.068	.825	1.983	8.932	-15.673	19.639

Next, a factorial analysis of covariance was used to assess whether information literacy scores were influenced by SES when treatment was taken into account. The ANCOVA was also used to visualize the data to see whether students with parents who have at least a college degree performed better than those who were the first in their families to attend college. For each dependent variable (posttest scores) the treatment was a significant factor, but the SES was not, as illustrated in the following sections.

Socioeconomic Status and Evidence of Research

A factorial analysis of covariance showed that for Evidence, the treatment made a significant difference, but the education level of parents did not. The effect of SES was not significant (F[1, 158] = .407, p =.05, η^2 = .003), but the treatment was significant (F[1, 158] = 19.094, p = .05, η^2 = .110). The interaction between SES and treatment was also not significant (F[1, 158] = .191, p = .05, η^2 = .001). The observed power for treatment is high, but low for statuses (see Table 5-42).

Table 5-42. Significance of Socioeconomic Status and Treatment on Evidence Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	103.135[a]	3	34.378	6.547	.000	.112	19.642	.969
Intercept	2853.896	1	2853.896	543.527	.000	.778	543.527	1.000
SES	2.137	1	2.137	.407	.524	.003	.407	.097
Treatment	100.254	1	100.254	19.094	.000	.110	19.094	.991
SES Treatment	1.001	1	1.001	.191	.663	.001	.191	.072
Error	813.858	155	5.251					
Total	4882.000	159						
Corrected Total	916.994	158						

a. R Squared = .112 (Adjusted R Squared = .095)
b. Computed using alpha = .05

The means for FIF (M = 3.652) and COLL students (M = 3.784) in the treatment group on the Evidence pretest at the beginning of the semester were similar to the means for the control group at the end of the semester: FIF (M = 3.583) and COLL students (M = 4.014). Yet the posttest means for both FIF and COLL students in the treatment group

were much higher than control group students in both socioeconomic groups. (See Table 5-43. and Figure 5-11).

Table 5-43. Means for Evidence by Socioeconomic Status for All Participants

		Evidence Pretest			Evidence Posttest		
SES	treatment group	Mean	Std. Deviation	N	Mean	Std. Deviation	N
FIF	control				3.583	1.6650	18
	treatment	3.652	1.9575	46	5.511	2.3933	46
	Total				4.969	2.3669	64
COLL	control				4.014	2.2046	35
	treatment	3.784	2.1174	58	5.592	2.4138	60
	Total			104	5.011	2.4495	95
Total	control				3.868	2.0314	53
	treatment				5.557	2.3938	106
	Total				4.994	2.4091	159

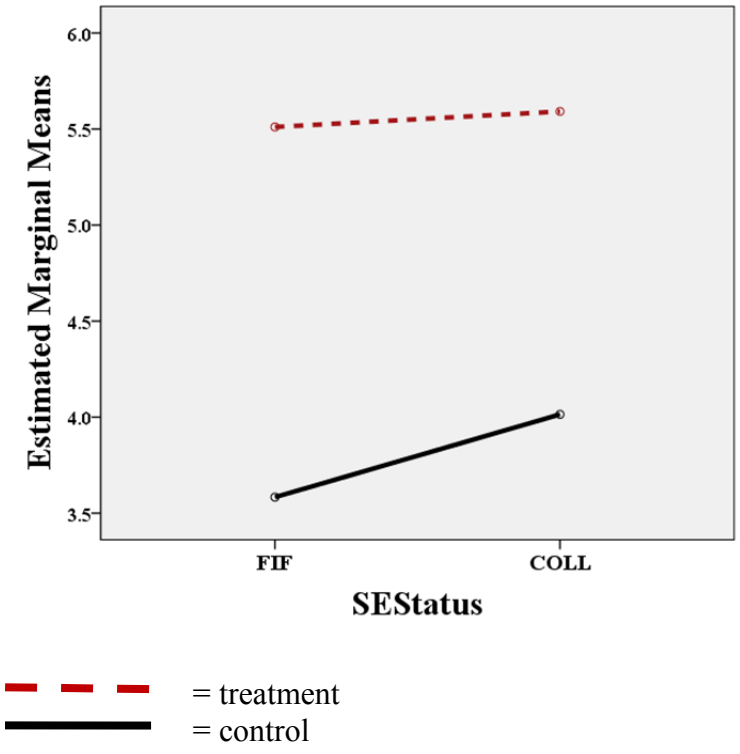


Figure 5-11. Comparison of Means for Evidence by Treatment Group and by Socioeconomic Status

A t-test failed to reveal a statistically reliable difference between the mean score for Evidence post-test between FIF and COLL students for all students in the study who took the post-test. There was no significant difference in the Evidence post-test score for FIF (M=4.97, SD=3.47) and COLL students (M=5.011, SD=2.44) conditions; t (157)=.-.107, p = 0.05. The difference in scores on Evidence between FIF and COLL students is within the margin of error as Table 5-11 illustrates.

The treatment made a significant difference in performance on the Evidence posttest, but the education level of parents did not. This result is consistent with the lack of significance found between socioeconomic groups with regard to GPA. FIF students in the treatment group started slightly lower than COLL students and gained more than COLL students when comparing control to treatment groups on the Evidence of Research criterion. This indicates that the treatment was strong and SES was not a factor. Indeed, students in the treatment group who were the first in their family to go to college did as well or better than students with at least one parent who had graduated from college in both treatment and control groups, and contradicts Hypothesis 3.2.

This is a major finding, because people often assume that students from better educated families have a head start and therefore usually continue to exhibit superior achievement. This may be true for high school, first and (perhaps) even second-year college students in other contexts, but it was not found to be true for the students in this study who were experienced upper-division undergraduates. Sirin's (2005) data is more than a decade old, and it is possible that the reduced effect of SES on achievement that he found has continued to trend down for other reasons as well. For this study in particular, context and demographics may also account for the unexpected results. In the case of

GPA, hard work might account for the lack of significance, especially at a public university where more affordable tuition and proximity to home attracts a high percentage of children from local working class and immigrant families seeking to improve their status and earning potential through educational improvement. With regard to information literacy skills, we might conclude that in such an environment practice indeed pays off, regardless of a prior condition, such as family statuses, over which students have no control. This has implications for practice and research, which will be discussed in chapter 6.

Socioeconomic Status and Citation

Similar results pertained for achievement on the Citation posttest as reported in Table 5-44. The treatment was significant, but the SES was not. When viewed by SES, mean scores on the pretest were significantly lower than on the posttest, but the gain for students in the treatment group was even greater than for the Evidence trait. Although the difference in scores on Citation between FIF students in the treatment and control groups is wider than for Evidence, a *t*-test failed to reveal a statistically reliable difference between the mean score for Citation between FIF and COLL students.

In their ability to cite their sources, students who were the first in their families to go to college in both treatment and control groups performed similarly, as Table 5-45 and Figure 5-12 illustrate. However, students in the control group who came from better educated families did much worse than students in the treatment group who had similar parental education levels. Thus, even students with parents who presumably had learned to cite their sources, did much more poorly on citing than their socioeconomic counterparts who were given instruction and more opportunities to practice citing. The

intervention clearly moderated the difference and enabled all students in the treatment group to perform well on citation, regardless of statuses, by the end of the term. It is possible that students from the lower SES group had had much less exposure to citing sources, and this may account for their greater gains after exposure to the instruction and practice. This makes sense because Citation is probably the most formulaic of the information literacy skills, and therefore, probably the easiest to learn.

Table 5-44. Means Citations Posttest

SEStatus	Mean	N	Std. Deviation
FIF	4.511	46	2.7070
COLL	4.931	58	2.5503
Total	4.745	104	2.6162

	SEStatus	N	Mean	Std. Deviation	Std. Error Mean
Citation posttest	FIF	64	5.906	2.6769	.3346
	COLL	95	5.426	2.8944	.2970

Independent Samples Test

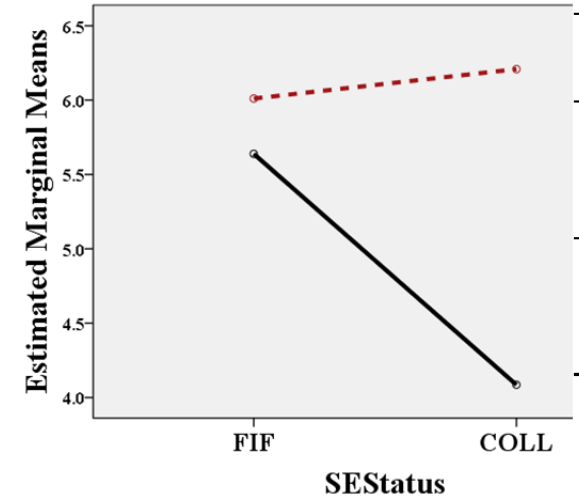
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Citation posttest	Equal variances assumed	1.692	.195	1.056	157	.292	.4799	.4543	- .4173	1.3772
	Equal variances not assumed			1.073	142.200	.285	.4799	.4474	-.4044	1.3643

Table 5-45. Comparison of Means for Citation Posttest by Treatment and by Socioeconomic Status

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	110.194[a]	3	36.731	5.005	.002	.088	15.015	.910
Intercept	3929.811	1	3929.811	535.474	.000	.776	535.474	1.000
SEStatus	15.000	1	15.000	2.044	.155	.013	2.044	.295
Treatment	50.787	1	50.787	6.920	.009	.043	6.920	.744
SEStatus * Treatment	25.012	1	25.012	3.408	.067	.022	3.408	.450
Error	1137.536	155	7.339					
Total	6268.750	159						
Corrected Total	1247.730	158						

a. R Squared = .088 (Adjusted R Squared = .071)

b. Computed using alpha = .05



Descriptive Statistics for Citation posttest

SEStatus	treatment group	Mean	Std. Deviation	N
FIF	control	5.639	2.9843	18
	treatment	6.011	2.5744	46
	Total	5.906	2.6769	64
COLL	control	4.086	2.8167	35
	treatment	6.208	2.6622	60
	Total	5.426	2.8944	95
Total	control	4.613	2.9412	53
	treatment	6.123	2.6139	106
	Total	5.619	2.8102	159

= treatment
= control

Figure 5-12 Comparison of Means for Citation Post-test by Treatment and Socio-economic Status

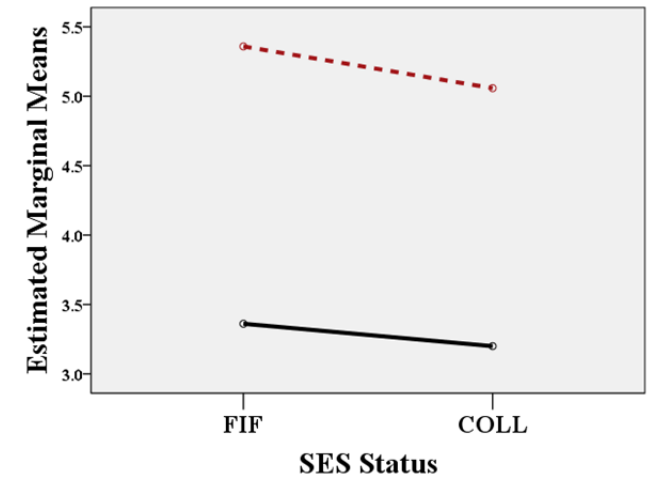
Socioeconomic Status and Integration

Results for student performance on the Integration posttest, that is, the ability of students to integrate their research and sources into their arguments, again showed that the treatment was significant, but that SES was not. A *t*-test failed to reveal a statistically reliable difference between the mean score for Integration between FIF and COLL students. The difference in scores on Integration between FIF and COLL students is within the margin of error (see Table 5-46 and Figure 5-13).

Table 5-46. Comparison of Means for Integration by Treatment Group and by Socioeconomic Status

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	132.993[a]	3	44.331	9.033	.000	.149
Intercept	2439.375	1	2439.375	497.047	.000	.762
EDUCADJ	.740	1	.740	.151	.698	.001
TREATMT	124.229	1	124.229	25.313	.000	.140
EDUCADJ * TREATMT	.346	1	.346	.071	.791	.000
Error	760.699	155	4.908			
Total	4176.750	159				
Corrected Total	893.692	158				

a. R Squared = .149 (Adjusted R Squared = .132)



Descriptive Statistics

Dependent Variable:Integration posttest

SEStatus	treatment group	Mean	Std. Deviation	N
FIF	control	3.361	1.4122	18
	treatment	5.359	2.5683	46
	Total	4.797	2.4636	64
COLL	control	3.200	1.8036	35
	treatment	5.058	2.3199	60
	Total	4.374	2.3166	95
Total	control	3.255	1.6688	53
	treatment	5.189	2.4235	106
	Total	4.544	2.3783	159

= treatment
= control

Figure 5-13 Comparison of Means for Integration by Treatment Posttest and Socioeconomic Status

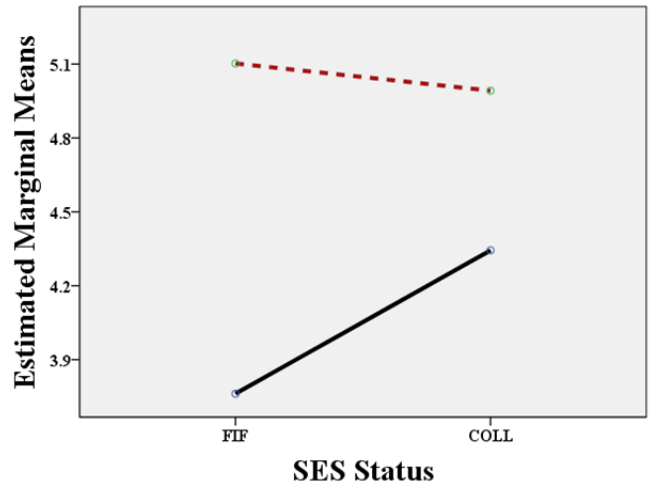
Socioeconomic Status and Writing

Similar results pertained for achievement on the Writing trait. The treatment was significant, but the SES was not. A *t*-test failed to reveal a statistically reliable difference between the mean score for Writing between FIF and COLL students (see Table 5-47). The difference in scores on Writing between FIF and COLL students is within the margin of error as shown in Figure 5-14.

Table 5-47. Comparison of Means for Writing Posttest by Treatment and Socioeconomic Status

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	34.807[a]	3	11.602	2.857	.039	.052
Intercept	2813.625	1	2813.625	692.720	.000	.817
SES	1.888	1	1.888	.465	.496	.003
Treatment	33.590	1	33.590	8.270	.005	.051
SES Treatment	4.070	1	4.070	1.002	.318	.006
Error	629.564	155	4.062			
Total	4221.000	159				
Corrected Total	664.371	158				

a. R Squared = .052 (Adjusted R Squared = .034)



Descriptive Statistics for Writing post-test

SES Status	treatment group	Mean	Std. Deviation	N
FIF	control	3.861	1.4431	18
	treatment	5.185	2.2245	46
	Total	4.813	2.1110	64
COLL	control	4.243	1.8165	35
	treatment	4.925	2.1009	60
	Total	4.674	2.0182	95
Total	control	4.113	1.6946	53
	treatment	5.038	2.1489	106
	Total	4.730	2.0506	159

Figure 5-14. Comparison of Means for Writing by Treatment and Socioeconomic Status

Socioeconomic Status and Holistic

Similar results pertained for achievement on the Holistic trait. The treatment was significant, but SES was not. A *t*-test failed to reveal a statistically reliable difference between the mean score for Holistic between FIF and COLL students. (see Table 5-48). The difference in scores on Holistic between FIF and COLL students is within the margin of error (see Figure 5-15).

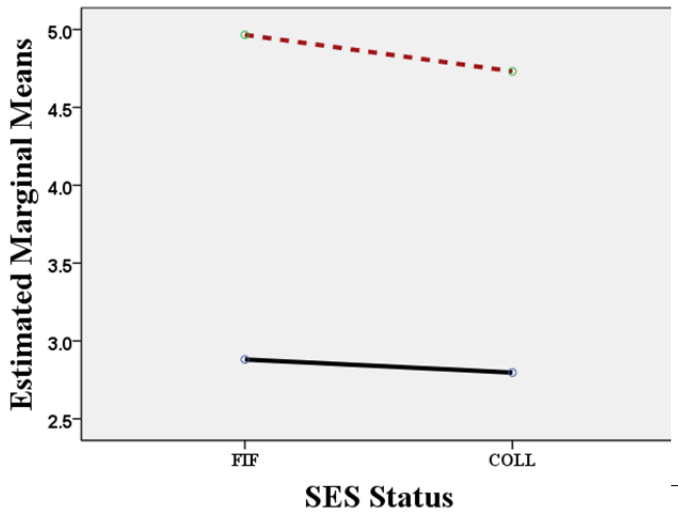
Table 5-48. Comparison of Means for Holistic by Treatment and Socioeconomic Status

Tests of Between-Subjects Effects

Dependent Variable:Holistic posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
7Corrected Model	142.638[a]	3	47.546	10.085	.000	.163
Intercept	2008.033	1	2008.033	425.936	.000	.733
EDUCADJ	.862	1	.862	.183	.670	.001
TREATMT	137.240	1	137.240	29.111	.000	.158
EDUCADJ * TREATMT	.192	1	.192	.041	.840	.000
Error	730.732	155	4.714			
Total	3628.373	159				
Corrected Total	873.370	158				

a. R Squared = .163 (Adjusted R Squared = .147)



Descriptive Statistics for Holistic post-test

SES Status	treatment group	Mean	Std. Deviation	N
FIF	control	3.0556	1.34917	18
	treatment	5.0652	2.37956	46
	Total	4.5000	2.31626	64
COLL	control	2.7143	1.86408	35
	treatment	4.6475	2.33767	60
	Total	3.9353	2.35917	95
Total	control	2.8302	1.70108	53
	treatment	4.8288	2.35385	106
	Total	4.1626	2.35110	159

Figure 5-15. Comparison of Means for Holistic by Treatment and Socioeconomic Status

5.9 Findings. Research Question 4: Other Factors.

What can we learn from how treated students use course materials and understand their own learning?

Hypothesis 4.1. Frequency of access of materials correlates with performance.

The online instructional materials relating to information literacy were made available to students in the treatment group via the course management system (Moodle). Activity is available from Moodle indicating every instance where a student clicked on any of the materials in the course management system for that course. This is a rich and voluminous source of data that turned out to require programming resources for analysis that were not available to the researcher. Nevertheless, the following analysis was carried out manually. Activity reports were exported from Moodle for several of the instructional resources prepared by the librarian to aid the students in their information literacy learning. Table 5-49 illustrates how many students used the librarian's materials in relation to their overall usage of the course materials. All students viewed resources in Moodle. The average number of views of all resources during the semester was 603 views. This is represented in the Table 5-49 as 'Mviews' and included any item in the course management system including viewing the syllabus, assignment instructions, online forum posts, students' profiles, etc. Three instructional resources were selected for closer analysis since they were the information literacy related materials most frequently accessed. 'Mcitations' represents a link to a research guide entitled *Understanding Citations* which explained proper citation, and provided links to websites that illustrate citation rules. 'Meval' and 'Mgoodpageex' represent narrated powerpoints entitled respectively *What Makes a Good Page, and Good Page Example,* that explain

the librarian's thinking when evaluating the quality of a Wikipedia page. These three resources were accessed at least once by 58 percent, 53 percent, and 84 percent of students, respectively. Only a few students accessed any of these individual resources more than once. The remainder of the information literacy related instructional materials were accessed by fewer than 50% of the students.

Table 5-49. Moodle Views Descriptive Statistics

	# of Students*	% of Students **	# of Students multiple accesses	Minimum # of views	Maximum # of views	Mean # of views	Std. Deviation
Mviews	106	100	N/A	171	4651	603.20	502.072
Mcitations	61	58	9	1	4	1.39	.803
Meval	56	53	17	1	5	1.85	1.236
Mgoodpageex	89	84	4	1	2	1.75	.500

n = 106

* = number of students who viewed the resource at least once

** = percentage of students in the study who viewed the resource at least once

An ANOVA was conducted to assess the relationship between frequency of access (Mviews) of online course materials and information literacy scores (see Table 5-50.) Similar results were found for the other measures of materials viewed (Mcitations, Meval, Mgoodpgeex). No significant relationships were found between use of information literacy materials and information literacy achievement. The three variables were analyzed in relation to the information literacy posttest scores for the treatment group. Few students consulted the materials prepared by the librarian and designed to aid their learning of the various information literacy skills. This is an important finding, because the students showed significant improvement despite their general lack of use of

the materials. If it was not due to the online instructional materials, it must have been the assignments themselves that made the difference.

Table 5-50. Effect of Student Views on Course Management System on Information Literacy Scores (Mviews)

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	Evidence posttest	16.180[a]	1	16.180	2.905	.091	.027	2.905	.393
	Citation posttest	20.998[d]	1	6.399	.922	.339	.009	.922	.158
	Integration posttest	20.121[e]	1	20.998	3.628	.060	.034	3.628	.471
	Writing posttest	21.408[f]	1	20.121	4.476	.037	.041	4.476	.554
	Holistic posttest		1	21.408	3.964	.049	.037	3.964	.505
Intercept	Evidence posttest	1107.196	1	1107.196	198.817	.000	.657	198.817	1.000
	Citation posttest	1478.533	1	1478.533	212.919	.000	.672	212.919	1.000
	Integration posttest	940.937	1	940.937	162.571	.000	.610	162.571	1.000
	Writing posttest	885.523	1	885.523	196.977	.000	.654	196.977	1.000
	Holistic posttest	796.063	1	796.063	147.395	.000	.586	147.395	1.000
Mviews	Evidence posttest	16.180	1	16.180	2.905	.091	.027	2.905	.393
	Citation posttest	6.399	1	6.399	.922	.339	.009	.922	.158
	Integration posttest	20.998	1	20.998	3.628	.060	.034	3.628	.471
	Writing posttest	20.121	1	20.121	4.476	.037	.041	4.476	.554

	Holistic posttest	21.408	1	21.408	3.964	.049	.037	3.964	.505
Error	Evidence posttest	579.169	104	5.569					
	Citation posttest	722.186	104	6.944					
	Integration posttest	601.936	104	5.788					
	Writing posttest	467.539	104	4.496					
	Holistic posttest	561.690	104	5.401					
Total	Evidence posttest	3846.000	106						
	Citation posttest	4739.000	106						
	Integration posttest	3497.500	106						
	Writing posttest	3198.000	106						
	Holistic posttest	3064.373	106						
Corrected Total	Evidence posttest	595.349	105						
	Citation posttest	728.585	105						
	Integration posttest	622.934	105						
	Writing posttest	487.660	105						
	Holistic posttest	583.098	105						

a. R Squared = .027 (Adjusted R Squared = .018)

b. Computed using alpha = .05

c. R Squared = .009 (Adjusted R Squared = -.001)

d. R Squared = .034 (Adjusted R Squared = .024)

e. R Squared = .041 (Adjusted R Squared = .032)

f. R Squared = .037 (Adjusted R Squared = .027)

Hypothesis 4.2. Students understanding of their own research/learning process correlates with their performance.

A sample of 20 reflective statements, including a range of pre- and posttest scores, were read, and an initial list of topics was developed to identify key elements of interest that surfaced in the student reflections. After a list of items had been created, the entire set of student reflections was scored by the researcher. Each category became a dichotomous variable indicating whether or not the student mentioned the trait. The categories were coded as follows.

Instructions was the label used to designate students who explicitly wrote that they had read the instructions.

Decided was used for those students who stated that they thought about the question, made up their minds, and then sought information that supported their conclusion. These were students who did the opposite of those whose reflective statements showed an authentic research process.

Authentic was used to indicate the students who described a research process in which their minds were open at the beginning and that they conducted research to understand the issues.

Articles was used to denote students who explicitly stated that they had sought and read the articles suggested in the assignment prompt.

Read was noted if the student mentioned having done a lot of searching and/or reading in order to fulfill the assignment.

Library was the indicator for students who mentioned using the library to find sources.

Alternatives was used for reflective statements that included the concept that alternative viewpoints were sought out or considered.

Selection Quality denoted students who discussed the activity and issue of selection and/or quality of sources as part of their process.

Finally, the term *Transformation* was used for students who acknowledged that they had made an error in their earlier work in the course and had come to a better level of understanding by the end of the term. It was only applicable to reflective statements written at the end of the semester by students in the treatment group.

At the pretest stage, 21% of students explained that they first sought information before deciding, that is, they did authentic research. Fifty-six percent stated that they decided first and then looked for materials that agreed with their point of view. The remainder did not comment either way. Over half the students (54%) wrote about giving attention to the quality of the sources. A relatively small percentage of students mentioned locating and reading the sources mentioned in the prompt (*Articles*, 15%), searching or reading a significant amount (*Read,* 15%), or using the library (*Library*, 14%).

Six of the nine process categories were mentioned more at the end of the semester than at the beginning: *Instructions, Authentic, Read, Library, Alternatives,* and *Selection/Quality.* Two categories, *Decided* and *Articles,* received less notice at the end of the semester than at the beginning. Fewer students wrote about 'deciding before writing' at the time of the posttest. This is logical because they probably would not have seen this as a relevant part of the process at that stage. However, many students began their posttest reflections by editing their pretest reflections, and this may account for the

still substantial number (37%) who mentioned deciding first. Twelve percent of the students in the treatment group who submitted a posttest and a reflective statement remarked on the fact that they had learned something. Although over 60% of the students improved their scores, only 12% of the students in the treatment group commented on their learning in their reflective essays at the end of the term as shown in. Table 5-51. See Figure 5-16 for a visual comparison of the details.

Table 5-51. Average GAIN Score Frequencies

	% Treated Students whose Info Literacy Scores were:		
	Worse	**Same**	**Better**
GAIN Evidence	9%	16%	**75%**
GAIN Citation	22%	17%	**61%**
GAIN Integration	15%	15%	**70%**
GAIN Writing	21%	14%	**65%**
GAIN Holistic	16%	15%	**69%**

This is interesting because the objective findings showed a much higher percentage of students had actually improved their scores on the posttest than commented on their gains in understanding in their reflective statements.

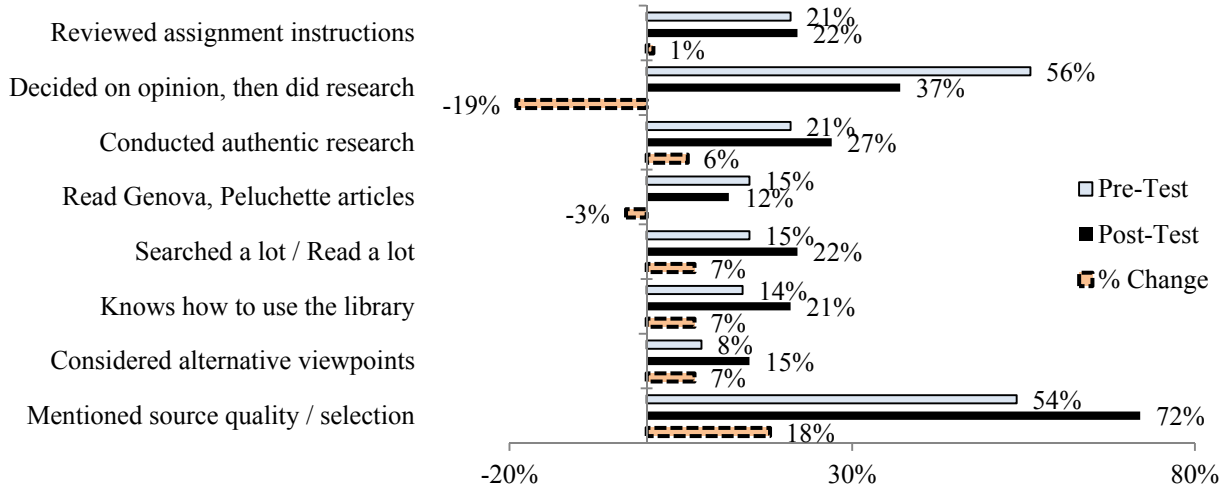


Figure 5-16. How Treated Students Reported on Their Process

Of all the relationships between the pretest reflection codes and their pretest performance, only "considered alternative views" correlated significantly with the pretest holistic scores. No other significant relationships were found. This may be explained by the theory that students who considered alternative viewpoints enough to comment on it as an important part of their process may have been more critical thinkers (as reflected in their overall scores) as shown in Table 5-52.

Table 5-52. The Impact of Alternative Viewpoints on the Holistic Score

Chi-Square Tests – Ealtview & Pretest Holistic Score

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	41.661[a]	11	.000
Likelihood Ratio	29.387	11	.002
Linear-by-Linear Association	4.606	1	.032
N of Valid Cases	87		

a. 16 cells (66.7%) have expected count less than 5. The minimum expected count is .09.

Small correlations were found between several of the variables as shown in Table 5-53. Students who mentioned reading the instructions were more likely to have said they read the articles in the prompt and had done much reading and searching. There was a small negative correlation between deciding in advance of doing the research, and reading the articles in the prompt. The highest correlation was found among students who were authentic researchers. They were more likely to have read the articles in the prompt, have done more reading and searching, and knew how to use the library. There were no significant relationships between *Decided/Authentic* and gender or transfer.

Table 5-53. Correlations for Reflective Statements on Pretest

		Instruct	Decided	Authentic	Articles	Read	Library	Alt Views	Quality
Instruct	Pearson	1	.043	.089	.263*	.263*	.025	-.163	.130
	Sig.		.694	.410	.014	.014	.820	.133	.232
Decided	Pearson		1	-.592**	-.223*	-.158	.011	.001	-.084
	Sig.			.000	.039	.147	.920	.993	.442
Authentic	Pearson			1	.343**	.263*	.025	.132	.016
	Sig				.001	.014	.820	.223	.885
	N				87	87	87	87	87
Articles	Pearson				1	.277*	.277*	.090	.128
	Sig.					.010	.010	.408	.238
Read	Pearson					1	.096	.090	.063
	Sig						.378	.408	.561
Library	Pearson						1	.201	.063
	Sig.							.061	.561
Alt Views	Pearson							1	-.026
	Sig.								.813
	N								87
Quality	Pearson								1
	Sig.								

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

n = 87

Hypothesis 4.2. Essay length correlates with performance.

It has been shown that longer essay length correlates with higher scores on a timed standardized test of writing (Burstein et al., 2004). In this study, the students were allowed to take as long as they chose to, and although there was a minimum length, no maximum was specified, and as expected, essay length was positively correlated with higher information literacy scores. On average, students in the treatment group increased the length of their submissions from pretest to posttest on both the essay and the reflective statements. In addition, a t-test revealed a significant relationship between pre- and posttest word counts on both the essay and the reflective statements as reported in Table 5-54. Thus, posttest essay length was proportional to pretest length. These findings suggest that most students purposely wrote more when revising their work regardless of how much they wrote at the beginning of the term.

Table 5-54. Differences in Essay and Reflective Statement Length Pretest to Posttest

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean	Correlation	Sig.
Pair 1	Word count essay pretest	490.41	103	156.630	15.433	0.512	.000
	Word count essay posttest	749.43	103	343.153	33.812		
Pair 2	Word count reflection pretest	148.17	103	107.621	10.604	0.508	.000
	Word count reflection posttest	264.09	103	286.301	28.210		

Paired Samples Test - Paired Differences

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Word count essay pretest - wordcount essay posttest	-259.019	295.408	29.107	-316.754	-201.285	-8.899	102	.000
Pair 2	Word count reflection pretest – word count reflection posttest	-115.922	249.449	24.579	-164.674	-67.170	-4.716	102	.000

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

There were significant correlations between each of the dependent variables and the amount of writing submitted by students in the treatment group, for both the essay pre- and posttests and in the pre- and post reflective statements, as reported in Table 5-55. The correlation between essay length on the pretest as compared to essay length on the posttest remained approximately the same for Citation and decreased slightly for Evidence, Integration, Writing and Holistic variables. Since citations have a fixed length and are independent of essay length, this finding is understandable. For the other criteria, those aspects of information literacy for which longer essays afford the opportunity to better explain and demonstrate information literacy behavior, this finding provides evidence that essay length was not a confounding factor, but a result of the treatment.

Table 5-55. Correlations between Dependent Variables and the Essay and Reflective Statement Length Pretest to Posttest

Correlations

		Evidence		Citation		Integration		Writing		Holistic	
		pre	post	pre	post	pre	post	pre	post	pre	post
wordcount essay pretest	Pearson Corr.	.538**	.409**	.320**	.328**	.537**	.390**	.452**	.421**	.466**	.399**
	Sig. (2-tailed)	.000	.000	.001	.001	.000	.000	.000	.000	.000	.000
	N	103	105	103	105	103	105	103	105	103	105
wordcount reflection pretest	Pearson Corr.	.475**	.267**	.279**	.204*	.337**	.256**	.288**	.334**	.232*	.279**
	Sig. (2-tailed)	.000	.006	.004	.037	.000	.008	.003	.000	.018	.004
	N	103	105	103	105	103	105	103	105	103	105
wordcount essay posttest	Pearson Corr.	.451**	.470**	.296**	.325**	.348**	.444**	.302**	.404**	.282**	.381**
	Sig. (2-tailed)	.000	.000	.002	.001	.000	.000	.002	.000	.004	.000
	N	104	106	104	106	104	106	104	106	104	106
wordcount reflection posttest	Pearson Corr.	.487**	.358**	.307**	.263**	.325**	.348**	.244*	.349**	.279**	.352**
	Sig. (2-tailed)	.000	.000	.002	.006	.001	.000	.013	.000	.004	.000
	N	104	106	104	106	104	106	104	106	104	106

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

There were small but significant correlations between essay wordcount and gain scores in Integration, Writing, and Holistic criteria as reported in Table 5-56. This further confirms the conclusion that information literacy instruction may result in students writing longer essays, which in turn results in better performance and more evidence for assessment. Thus, increased essay length may be both an indicator of information literacy, and an instructional goal.

Table 5-56. Correlations between Word Count and Gain Scores

Correlations

		Evidence gain score	Citation gain score	Integration gain score	Writing gain score	Holistic gain score
Word count essay posttest	Pearson Correlation	.088	.047	.256**	.233*	.269**
	Sig. (2-tailed)	.376	.638	.009	.017	.006
	N	104	104	104	104	104
Word count reflection posttest	Pearson Correlation	-.059	-.048	.126	.197*	.206*
	Sig. (2-tailed)	.554	.631	.203	.045	.036
	N	104	104	104	104	104

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Course Grades

General level of achievement thus far in college as measured by GPA at the start of the semester under study correlates well with information literacy scores as indicated in Table 5-57. This is expected.

Table 5-57. GPA Correlations with Information Literacy Variables

		E1evi adjtotal	E2cit adjtotal	E3int adjtotal	E4writ adjtotal	E5hol adjtotal	T1evi adjtotal	T2cit adjtotal	T3int adjtotal	T4writ adjtotal	T5hol adjtotal
GPA	Pearson	.404**	.264**	.418**	.412**	.423**	.494**	.419**	.502**	.447**	.458**
	Sig.*	.000	.006	.000	.000	.000	.000	.000	.000	.000	.000

*Sig. (2-tailed)

There were no correlations between SAT Verbal Critical Reading and the information literacy variables.

CHAPTER 6: CONCLUSIONS

6.1 Introduction

The results of this research show that a diagnostic research essay can be used to measure and aid information literacy learning in a sample group of upper-division college students at a technology university. The researcher used an experimental research design and multivariate statistical methods of analysis to examine contextualized information literacy in authentic student work products and explored factors that were thought to affect student performance. There were five major findings:

1. The method used for measuring information literacy was shown to be statistically reliable and valid.
2. The use of the brief essay as a pre- and posttest showed that the students in the treated group achieved statistically significant gains in higher-order skills associated with information literacy.
3. The students in the treatment group significantly outperformed students in the control group with substantive effect sizes.
4. Socioeconomic status had no significant impact on information literacy performance.
5. Student use of online instructional materials had no significant impact on information literacy.

Conclusions and limitations are discussed in this chapter. In addition, this work will be viewed in relationship to other research with a proposal for the types of follow-up study needed to continue this line of research. This study added to the small body of research that has used empirical evidence from authentic student work to assess information literacy. The method enabled the researcher to partition the influence of multiple

variables and their effect sizes. This lays a foundation for future research where systematic measurement of predictor variables can be used to understand which variables are important and most effective in improving learning outcomes. Furthermore, this study is unique in its quantitative assessment of ACRL Standard Four, information use, and students' ability to integrate sources into written arguments. A summary of findings and interpretations is reported in Table 6-1.

.

Table 6-1 Summary of Findings

Research Question &Hypothesis	Concept	Finding	Interpretation, Relationship to Other Research, Implications for Future Research
RQ1. Is the brief essay a reliable and valid instrument for assessing higher-order information literacy skills?	Instrument & Process		
H 1.1. The brief essay can be reliably scored.	Reliability	Cronbach's alpha is high. The scores of multiple readers are in sufficient agreement so that the scoring method may be considered reliable.	Well trained readers using clearly defined criteria can come together in agreement even using a short essay.
H 1.2. The criterion variables used will be internally consistent.	Reliability	Interitem correlation is high. The five measures of information literacy have a good level of consistency among items.	The instrument is a meaningful one. Merging of writing and research criteria was new. This has been mentioned by many, but this is the first data found to illustrate it.
H 1.3. The brief essay (operationalization) is a valid instrument for assessing higher-order information literacy skills.	Validity	• Face validity-professionals agree it makes sense; • Essay length matters. • Content validity-mappings to ACRL + Bloom • GPA correlates to IL scores • High/Low indexes logically related	Longer essay may impede ease of scoring, but encouraging slightly longer essay may yield more insight. Some shortcomings were found in the use of a very short essay. Students were asked for their 'feelings' in the prompt. This wording was an inadequate pre-operationalization of the instrument that resulted in a threat to validity. Alter question wording in future studies.

Research Question & Hypothesis	Concept	Finding	Interpretation, Relationship to Other Research, Implications for Future Research
RQ2. Will the intervention designed by a collaborating researcher and instructor improve students' ability to use researched information effectively in their written work?	Value of Instruction		
H 2.1. There will be a statistically significant improvement in student performance on the research essay following the intervention specifically targeted at developing the component higher-order information literacy skills.	Improvement in Info Literacy	Paired sample t-test showed treated students showed a significant improvement on average. Repeated Measures ANOVA showed improvement, with moderate effect sizes. ANCOVA showed improvement even when effect of pretest was removed. Compare pretest and GAIN scores. Diagnostic traits correlate with GAIN scores except for holistic. Evidence of Research was the measure most improved, and Writing least improved.	Students showed a degree of improvement in one course greater than that shown by many students after 4 years of college. Students benefited overall from the instruction and assignments provided throughout the semester. Targeting component information literacy skills with specialized instruction may indeed improve them. The intervention worked, but we cannot isolate the single importance of particular variables that influenced the outcome. Alignment of objectives, learning activities, and assessment may exert a major influence on the outcome. "A strong relationship exists between the quality of work a student produces upon entering a course with work produced by that student at the end of the term." This diagnostic appears useful. This is similar to findings in the writing portfolios. Citation was usually found to be the strongest correlate. This provides some confirmation for those studies using bibliographies as proxies for information literacy.
H2.2. Students in the treatment group will perform significantly better than those in the control group.	Treatment vs. Control	Students in the treatment group performed significantly better than those in the control group. Moderate effect size. One-Way ANOVA had similar results, slightly smaller effect sizes. When skills grouped into low/high skills, the means for the holistic score for the treatment group (M=4.818) and the control group (M=2.830) are both lower than the means for the low skills and the high skills indexes.	The embedded librarian made a difference, but it was not the instruction, because H4.1 results showed students did not use the instructional materials. The assignments embodied the instruction. This indicates the holistic score is, as it was meant to be, independent and not a summary of the other skills.

Research Question & Hypothesis	Concept	Finding	Interpretation, Relationship to Other Research, Implications for Future Research
RQ3. Which prior conditions had a significant effect on a student's information literacy performance?	Conditions Affecting Performance		
H 3.1. The pretest accounts for a portion of the variance in scores between the pre and posttest.	Conditions Affecting Performance	Excluding the effect of the pretest, students still performed better than those in the control group. 30% of the variability is explained by the treatment after removing any influence of the pretests.	Simply taking the pretest had a positive effect on students in the treatment group. The pretest is a not only a good diagnostic, but a good instructional strategy. Practice helps. It also supports constructive alignment and makes students aware of course objectives and assessment criteria at the beginning of this process.
H 3.2. SES will show a significant correlation with information literacy scores with a meaningful effect size.	Conditions Affecting Performance	SES as measured by highest level of parent's education had no measurable impact on information literacy performance. Nor had it been correlated with GPA at the beginning of the term.	Because these were upper-division undergraduates, perhaps two years of college smoothed these results. (This was tested on first year students in Fall 2012). Perhaps the intervention was strong enough to overcome these initial differences.

Research Question & Hypothesis	Concept	Finding	Interpretation, Relationship to Other Research, Implications for Future Research
RQ4. What can we learn from student behaviors during the course?	Conditions Affecting Performance		
H 4.1. Frequency of access of materials correlates with performance.	Conditions Affecting Performance	No relationship between Moodle views and DVs. Observed activity levels lower than expected. Many students did not look at the materials at all.	If students did not learn from using instructional materials, they may have improved by simply doing the info literacy assignments. Educational data mining from course management systems is an avenue for future research.
H4.2. Student understanding of their own research/learning process correlates with their performance.	Conditions affecting performance	No significant correlation between reflection codes or additive index and performance. 56% students decided first / then researched.	Highlights the difference between surface and deep learning. Although this was an 'authentic' task, the students were not asked to do 'authentic' research. This is a distinction that has not been made before in the IL community. It is masked by discussion of low skills/high skills. Skills can be done at a low or high level, including the overall approach to research. More versions of the prompt should be tested to examine the relationship between IL and critical thinking.
H4.3 Essay length correlates with performance.	Conditions affecting performance	Both essay word count and reflection word count correlate with scores, essay more so. Posttest responses were longer.	On the one hand this result makes the pretest a good diagnostic tool. Posttest may be more questionable as students are more savvy about how to improve their grades. If students learned they had to write more to explain better, no need to 'correct' research design for essay length because it is in itself an indicator of learning.

6.2 Reliability and Validity

In order to establish that the brief essay is a reliable and valid instrument for assessing higher order information literacy skills, three areas were investigated: the reliability of the scoring method, the internal consistency of the information literacy criteria, and the validity of the diagnostic. The high rate of interrater reliability showed that trained readers were able to distinguish among the criteria and consistently evaluate levels of performance. This has been demonstrated before, both in literature on writing assessments and in an earlier study (Scharf et al., 2007). However, successful replication of these results using the same methodology would be a welcome finding in librarianship, where replication studies are rare.

This study used methods for maintaining reliability that had been tested in prior research using direct assessment of writing portfolios (Scharf et al., 2007). Comparable reliability was achieved in this study, but there were important differences between the two studies. One aim of this study was to simplify administration while maintaining reliability. Previously, the method had been used on researched work of much longer length (i.e., writing portfolios), and on longer research papers where the students had the opportunity to write more, and thus the reader would have a better chance of observing desired outcomes. A limitation of the "assessment with rubric" method that has been frequently noted is its resource intensity. There is no doubt that such an assessment, even when streamlined, requires more time and intellectual investment than a limited response test. To lessen the amount of time needed , the required essay was reduced in length, and the number of assessment criteria was reduced from the earlier study. In this study, there were a total of five criteria, whereas the earlier study used two sets of five criteria each,

five for writing and five for information literacy. The amount of material evaluated was reduced from a portfolio (containing both researched and nonresearched writing) to a single brief researched essay. Another difference between the earlier study and the present one is that course instructors and collaborating librarians were used as readers in the earlier study, whereas the readers were drawn from outside the course and the institution in this one. This provides further evidence that people unfamiliar with the specific course can be trained as effective and consistent readers. Thus, the ability of the evaluators to consistently distinguish outcomes even in this brief writing sample is noteworthy. Given the high degree of reliability, the reduction in essay length as well as the reduction of the number of criteria has significant implications for practitioners, because it takes less time for raters to read less material and score it on half as many criteria.

It should also be noted that, unlike standardized tests of writing, this was not a timed assignment. Some students took more time than others and the essays ranged from one paragraph to over 10 pages. However, the typical paper of 2–3 double-spaced pages was adequate to the task. This was one element that constituted the authentic task. The implication for practice is clear. The brevity of the essay need not impede the reader's ability to assess the quality of the information literacy skills of the writer. Length is one among many of the choices a writer makes in crafting a written communication. In summary, this study achieved a high rate of interrater reliability in the reading of constructed responses despite the reduction in essay length and staff time needed for scoring. The writing prompt and the criteria used for assessment integrated the writing, critical thinking, and research skills. The criteria were internally consistent, yet readers

were able to distinguish among the traits and score reliably. This indicates that readers can discriminate between levels of achievement even when the text is fairly short, provided that the researchers have been adequately trained and employ clear criteria. Thus, the method has practical advantages, making it a viable method of authentic assessment for practitioners.

In the earlier study, a random sample of approximately 100 students was used to represent a larger cohort of students taking a particular course. This study had a similar number of participants, but used the work of all students who agreed to participate, rather than a sample. In both cases, the sample was representative of the demographics of the institution's student body. The study results were not intended to be generalizable beyond this particular institution. Rather, the study demonstrated that the model used can provide a structure for both local research and practice design. It demonstrated the value of aligning objectives, instructional planning, and assessment as a method for improving information literacy among college students.

The study performed well both in terms of translation and criterion validity (see Figure 4.4) (Trochim, 2001). The subjective evaluation of whether the criteria fit the construct of information literacy was validated in two ways. First, the validity of this study relies primarily on the underlying information literacy construct as articulated in the ACRL Standards, as well as in Bloom's Taxonomy of Educational Objectives. The mapping of dependent variables to these models of academic achievement provides evidence of the validity of the construct used in this study. (See Appendix E). The various elements of the diagnostic essay question required students to perform all the skills up to and including the highest levels depicted in each model. Second, this was

corroborated by educators at NJIT and by other academic librarians in New Jersey who were shown the diagnostic essay and found that it seemed to be a reasonable, integrated test of researched writing. Another measure of validity is the correlation of this measure of academic achievement with other measures of academic achievement. In this case, GPA correlated significantly with information literacy variables. This makes sense in this institutional context where the instructors involved have worked together previously and performed course-wide assessments that resulted in normed grading across a variety of assignments and students. Further studies may attempt to test the external validity of the diagnostic research essay on groups of students with similar—or markedly different—demographic characteristics.

6.3 Improvement and Engagement

The assignment was an authentic task that addressed all the ACRL Standards. The findings show, on average, a significant positive change in the quality of researched writing produced by students in the treatment group by the end of the semester. Moreover, on average, these students scored higher than the control group, even after partialling out the effect of the extra exposure and practice they gained in simply taking the pretest. The size of the effect is large in the context of other studies of student learning. In a cross-institutional study that used the Collegiate Learning Assessment to measure critical thinking, complex reasoning, and writing, students tested during their freshman year had improved by only 0.18 standard deviations three semesters later when tested at the end of their sophomore year (Arum & Roksa, 2011). Their results corroborated the meta-analysis done by Pascarella and Terenzini (2005) who reported similar results for studies done in the 1990s. They reported 0.50 standard deviation

overall from freshmen to seniors or 0.06 per semester for eight semesters. The mean improvement for treated students in this study was 1.45 standard deviations in one semester after averaging the gains for all the traits. The magnitude of this gain is striking, yet student use of the materials prepared by the librarian for online access was much lower than expected: Most students never looked at this material. They might have improved more if they had used the materials. This observation requires subsequent study. If the use of online materials does not have a clear impact on learning, it would suggest that librarians should spend less time on pathfinders and tutorials, and more on setting and communicating course objectives and assessment criteria. Indeed, the fact that many students showed significant improvement without using the instructional materials suggests that "try, fail, try again" is a good model. Like an athlete who can show dramatic improvement with clear goals, good coaching, and sustained practice, the student starved of research practice may make dramatic gains with clear learning objectives, the expert intervention of a librarian, and more practice.

Of course, the small size of this sample may have exaggerated the effect, yet the results may also be an accurate reflection of several factors. When the effect of taking the pretest was removed from the final result, the average improvement remained significant, even if the effect size was somewhat diminished. Something happened during the semester that made the difference. Because the treatment was at least partially responsible, and the positive effect of the pretest was also significant, one may conclude that giving a pretest, or administering the same assignment twice, may be a good instructional strategy. Replication and explorations of the types and frequency of practice are avenues for further study.

A student takes many courses while in college, but only some reward learning that explicitly includes critical thinking, reasoning, and writing like this course and this assignment in particular. Many institutions have introduced separate courses in critical thinking or learning to learn; some colleges include this focus in freshmen seminars or other general education courses. The results from this study strongly suggest that courses that embed information literacy as a form of critical thinking can yield positive results as well. The students in the treatment group were afforded multiple opportunities to flex their "research muscles." The individual tasks that comprised the instructional strategy were not isolated as targets of this study, but were considered as a whole. In addition to the diagnostic essay, there were several tasks that focused on writing, word use, definitions, and critical thinking that did not specifically target information literacy skills, yet were also intended to prepare the student for integrated writing/information literacy tasks.

Because it is not possible to discern from this study which aspects of the intervention were responsible, or to what degree, the individual assignments that operationalized the approach are good candidates for further exploration. The two information literacy tasks that led up to the WISER Project focused on skill building in finding, citing, and evaluating sources. First, the Green Web Page Evaluation task asked students to evaluate and compare three Web sites. After submitting their assignment, the students were able to view an expert analytical answer that had been created by a librarian. Second, the Wikipedia Source Evaluation task asked them to locate and review several references from a specific Wikipedia article on coaxial cable and evaluate them for format, quality, and relevance. The students were asked to verify whether or not the

reference actually supported the facts stated in Wikipedia by locating the original and comparing the two. After submitting their assignment, the students were again able to view a librarian's expert answer. This approach, where the students actively work the assignment and review the expert model only after submission, may be partially responsible for the improvement in performance.

The research design was intended to foster engagement as well as present a strategy for a successful intervention. The diagnostic essay itself was engaging—many students have said so, although the degree of engagement was not measured in this study. In their reflective essays, many wrote that they had already thought about the question of privacy on the Internet and had strong feelings about it. The level of engagement probably helped evoke writing illustrative of the information literacy criteria, which in turn enabled readers to observe the targeted behaviors. Using an assignment that elicited a high degree of engagement was important in order to obtain an honest effort from the students. However, the high level of engagement had an unexpected negative effect. In many cases, the students' strong feelings overwhelmed their logic even though students had been specifically instructed to consult outside sources as part of the assignment. Thus, the writing prompt might have been more effective if it had evoked less passion, since reasoned researched writing was the goal. Future studies might strike a better balance between an engagement that raises questions that in turn ignite exploration, and an engagement that evokes a passion that overwhelms logic. Also, a future study could use several different simultaneous prompts to test this understanding. The prompts should make it clear to students when logic and evidence are required rather than opinion. The WISER Project was also a source of motivation, because students were not researching

and writing for the instructors, but for the world. Students saw it as a piece of work they could proudly point to on the open Web and some began to understand themselves as independent actors in a community of Wikipedia editors rather than as merely students handing in required compositions.

Rubric Use and Self-Reflection

The use of a rubric for information literacy probably played a significant role in the degree of improvement. Rubrics have long been found to be helpful in writing assessment, and rubric use is among the recommended strategies for information literacy assessments. Oakleaf (2007) outlined four major barriers to the effective use of rubrics by librarians, which she suggested could be resolved by training. This study conformed to that model by using an outcomes-based approach, both analytic and holistic criteria, and providing effective communication about the criteria with both students and readers. Thus, her suggestion that barriers to rubric use could be addressed through training were confirmed by this study.

The use of rubrics in information literacy assessment has been the subject of much discussion, but with limited study. The first published study documenting the use of a rubric for evaluating information literacy also examined the relevance, quality, and use of sources chosen (Emmons & Martin, 2002). This investigation was similar in many respects to the Emmons and Martin's study, but added a holistic score as a variable. Emmons found that information literacy instruction by a librarian and the use of a researched essay slightly, but significantly, increased the relevance of sources used; however, the degree of improvement in source use and the integration of ideas were non-significant. What made the difference? Emmons and Martin studied research essays by

223 freshmen who were asked "to make and support an argument using sources other than the assigned course readings." The skill needed to identify a researchable topic based on the course content is a high-level skill, yet it was the first step to fulfilling Emmons and Martin's assignment. In the present study of 105 upper-division students, the topic and a timely question were provided, and the prompt alluded to sources that were incomplete but findable. Thus, the hard work of selecting a researchable topic of interest was eliminated as a variable.

Students showed a high level of concern about the sources they used in their reflective statements. Perhaps by saving time usually spent on topic selection, students were able to focus on other information literacy activities, which resulted in improved outcomes in the areas measured. The students commented more on source selection and quality than any other topics in their reflective statements. Over 50% discussed source selection and quality in the pretest reflection, and that percentage increased by 18% in the posttest. This corresponds to the actual improvement on the Evidence of Research criterion, which showed the largest jump in average scores from pretest to posttest. This degree of improvement may well be due to the emphasis placed on source selection and evaluation throughout the course in grading rubrics, and by the multiple opportunities to practice.

In Emmons and Martin's (2002) study, the authors concluded that despite the gains, they were not satisfied that students were actually finding sources relevant to their topics, because the students' treatment of the topics was often superficial. In subsequent qualitative studies, in which a small number of student papers were examined closely, Emmons, Martin, Botts, and Amundson (2010a, 2010b) confirmed their suspicion that

students' use of sources was usually shallow. Emmons et al. concluded that students continued to view the researched essay as an academic exercise rather than as a quest for knowledge. Despite this study's more impressive gains in several areas, the absolute scores and the reflective statements of students also revealed the relatively shallow nature of student inquiry. The posttest averages ranged from 4.8 to 6.1 on a 10-point scale. Although the improvement is impressive, the final scores are still mediocre. Reflective statements showed that a majority of the students were not seeking an answer, or exploring a topic to learn more about it, but only seeking instances of agreement with their own opinions. Students also wrote rather short reflective essays, without much depth, so it seems they were not particularly self-aware or able to articulate their process. Many students did not mention some of the activities the researcher thought might be important to improving their information literacy performance. For example, fewer than 25% of the students commented on reviewing instructions, reading and searching, using the library, or considering alternative viewpoints. Yet the number of students who commented on these topics increased by the end of the term. Thus, the positive change indicates that more students used these methods following the intervention, or at least realized they were important.

In a related finding, only 12% of the students in the treatment group acknowledged gains in learning, despite the evidence that a much larger percentage did indeed improve. This discrepancy may be explained as follows. First, it is possible that the students may have gone through many of the steps in the research process but had not written them down, either because they were not consciously aware of the steps they had taken, or because they were not compelled to provide details in writing. Alternatively,

these processes might not be as necessary as the researcher hypothesized. Perhaps simply being asked to reflect on the process had the largest impact, given that it has been shown that self-reflection aids learning. Further study is needed to isolate and explore this interpretation. The reflective statements are self-reported data that could be prone to self-censorship for a variety of reasons. NJIT STEM students, who predominate in this course, are rarely asked for written self-reflection. Perhaps it takes practice to become more self-reflective, or at least more adept at articulating a process. Perhaps one semester is not enough time to realize one has learned something. This is another good avenue for a follow-up study to further test the relationships among self-efficacy, writing, and information literacy.

6.4 Confirmation-Seeking Behavior

In the pretest reflective statement, 21% of the students described a method that could be called "authentic research." That is, they explored the topic before coming to a conclusion. Some (23%) did not comment on this at all. However, a majority (56%) of students explicitly described their method as one that was essentially the opposite of conducting authentic research. One student's reflective statement illustrates the method described by the majority of students. This behavior might be summarized in the language of the students as "I decided first, then I found sources that agreed with me." The following excerpt is typical of many. As one student wrote:

> In completing the initial task I first attempted to understand exactly what the given question was asking. In order to determine my opinion I speculated how I would react if my employer fired me because of pictures or comments I posted online. I then chose my position on the topic and

> began to research information that would help strengthen the validity of my argument. When a fitting article was found I analyzed the source in order to determine whether the information they provided was reliable. In this case both *The New York Times* and the CBS Evening News were reputable sources with great reputations for providing highly reliable information. The authors for both articles that I chose had extensive backgrounds with an abundance of published articles and editorials. There was no evidence that the information they provided was false or misleading and so I determined each source to be acceptable. The initial task was then completed using quotes from both articles to help create an effective argument for my position.

Another student put it more succinctly: "I first wrote down on a separate sheet of paper ideas based off of my views. Then I just free wrote my entire paper and went back and made corrections. Later I added in my sources and re-read my paper for mistakes." "Freewriting" is a stream-of-conscious method of writing typically used as a way to get started when a writer is blocked. This technique is sometimes employed when a writer is "stuck," but this usually comes at the writing stage, rather than prior to conducting research, as this student did. In doing so, the student avoids any authentic research or opportunity to consider alternative viewpoints. In reflecting on source selection, another student wrote: "I narrowed in on two sites that supported what I believed in." This method was expressed by many students.

We can compare this with self-reported data on students' research habits from Head's Project Information Literacy (PIL) study (Head & Eisenberg, 2010). PIL study

students reported being most concerned about finishing the work and getting a good grade, a finding compatible with comments from participants in this study. However, three-quarters of the students in the PIL study reported that doing authentic research and learning were important to them. This belief was not supported by actions and reflections observed in about 80% of the students in this study, who did not conduct authentic research or mention learning in their reflective statement. However, the response to a different question in the PIL study found that three-quarters of the students thought “it was also important to find answers to insert in their paper to prove to the instructor the research part of the assignment had been done.” This fits with the finding that the most frequent approach taken by students was to find sources that agreed with them as a means of fulfilling the research portion of the assignment. The need to “find answers” to an assignment that poses an analytical question with no right answer highlights a fundamental difference in student and instructor objectives for a research-based essay. Therefore, a closer alignment of instructor and student objectives may further boost learning.

This assignment was crafted so as to encourage a high degree of student engagement, but in doing so, it was justifiably interpreted by students as asking primarily for their opinion. Thus, deciding first and then seeking confirmation in sources seemed like a logical process. However, this "confirmation-seeking" behavior, observed by many academic librarians, but so far unnamed, differs from hypothesis confirmation in scientific or in public library reference contexts. There is little attempt at open-minded discovery, which might lead to hypothesis disconfirmation. It also differs from the “imposed query (Melissa Gross, 1995), a common situation among college assignments

that was explored in the literature on assigned tasks in information retrieval scholarship (Limberg & Sundin, 2006; Tanni & Sormunen, 2008). This type of confirmation-seeking behavior also describes a situation in which a person is seeking verification of a known or supposed fact, but is open to a negative result. In this case, many participants sought only confirmation of their own opinions and were biased in their selection against contradictory information. Because the students were asked to make a judgment, and it was a question relevant to their own lives, many allowed their initial emotional response to lead their research, rather than beginning with an exploration of the issues via the secondary literature. (A minority of students explored the topic before taking a position.) The majority's approach has been documented in psychological research concerning the competition raging in our minds between emotion and logic (Kahneman, 2010). The method used by these students—and many others—has long been explained by philosophers and psychologists as "confirmation bias," that is, the tendency to seek and find information that confirms one's beliefs. A companion theory, "selective exposure," suggests people both seek confirming information and avoid contradictory evidence (Nickerson, 1998). Zipf's "principle of least effort" (1949) also seemed to be operating, because many students took the most efficient path to completing the assignment by avoiding genuine research, which would, of course, take more effort.

These behavioral theories can help to provide an understanding why most students have frequently taken this path. Several seminal theories in library and information science that describe question formulation can also be brought to bear. The first is Taylor's information needs (1962); the second is Belkin's anomalous state of knowledge (1980); and a third is Dervin's sense-making (Dervin & Nilan, 1986). All

three theorists addressed the state of a researcher's mind prior to engaging in authentic research. People seek information because they have a need whether it is unfocused, unrecognized, or the result of a gap. Yet here is a condition where the student seeks confirmation to support an imposed need. Some information scientists have begun to tackle assigned learning tasks as a class of "imposed query" (Gross, 1998). In a review of research on information behavior in assigned learning tasks, Tanni and Sormunen (2008) report on several studies that indicate that students often turn research tasks into reporting tasks, and focus on the final product rather than the learning. The student behavior observed in this study fits those findings, and that of others who have questioned, like Emmons, the depth of cognitive processing that is taking place in fulfilling assigned researched writing tasks (Emmons & Martin, 2002; Emmons, Martin, Botts, Amundson, 2010a, 2010b). It brings to mind the work of Marton and Saljo (1976), who, through an empirical study of Swedish post-secondary students, identified the difference between surface-level and deep-level information processing. More recent research coming out of the Citation Project (Howard et al., 2010) has shown that many students piece together snippets of text from various sources using surface-level processing, without engaging at a level deep enough to gain an understanding that would lead to an interpretation that could ultimately be expressed in their own voices.

Many students appeared to have views of what research entails that differed from the understanding of those charged with helping them learn. This mismatch between faculty expectations and the information behaviors of students, the majority of whom are digital natives, may be the result of different assumptions and frameworks for understanding scholarly communication. These disconnections are frequently manifested

in emotional terms, such as when a professor laments the lack of engagement, or a student laments the lack of relevance of the assignment. The use of Wikipedia was effective in bridging this gap. For the instructor, it required that students employ scholarly conventions of documentation. For students, it was a chance to engage in an online community around a topic of personal interest. This has implications for how we teach, and what we teach and assess. The mechanics of research are commonly taught in one-shot sessions and in online tutorials. This finding provides a clue that information literacy skills might be improved if librarians insist on continuous involvement in authentic research, research that requires a mind being open to an unexpected answer. In addition, authentic research may be a casualty of, for example, an instructor's requirement for a "thesis statement" early in the research process. This is a hypothesis that could be tested in future research studies, and it might yield results with significant implications for practice if it provides evidence that the sequencing of tasks matters. Confirmation-seeking behavior may be assessed as part of ACRL Standard One, "[t]he student knows when information is needed." This study did not include a direct measure of Standard One, but the discovery of confirmation-seeking behavior through analysis of the reflective statements highlights a phenomenon that merits further attention. Researchers and practitioners may wish to explore further the effect of alternative prompts on student behavior concerning information need.

Learning came through a problem-solving approach with students discovering the parameters and protocols of the community and audience they were addressing. It motivated students at first because it seemed "cool," and later because they realized their entry, if well done, could remain out on the open Web for all to see. It changed their

perception of the assignment from the static class assignment to a dynamic, content-rich writing activity that used a topic about which they cared deeply and that they could publicly author (or that they could improve an authoritative article). It heightened their awareness of the difference between facts and opinion and to the true purpose of citation, that is, to validate facts using reliable sources. It covered every ACRL standard with little explicit direction. Wikis and Wikipedia are changing and evolving, so the conditions could change. Nevertheless, Wikipedia can be used to create an occasion for engaging students in active-learning about integrated research and writing skills.

In summary, although students improved their ability to use sources to support a point of view, most frequently students completed the assignment using a shortcut technique of foraging for confirming sources rather than conducting authentic secondary research. However, the fact that a quarter of the students interpreted the assignment as a strong situation requiring authentic research without explicit prompting to do so is promising. It would be useful to examine student work to explore further under what conditions students use authentic research techniques, as well as deep- versus surface-level processing of information. Altering the language of this assignment, and experimenting with different questions that would privilege authentic research may yield different results. In addition, a question that requires the use of more sophisticated sources and research techniques may help instructors focus student learning on a more critical reading of the assignment and deeper understanding of research skills and applications.

6.5 Additional Considerations

Hypotheses about the effect of prior conditions on information literacy achievement were not borne out by this study. The most surprising finding was the lack of a correlation between socioeconomic status and performance. This may have been due to the fact that the study participants were upper-division undergraduates in a technical university, thus two or more years at college had smoothed out the differences that may have been evident when they were freshmen. Another surprising finding was the lack of a correlation between transfer status and performance. The academic "street wisdom" at this institution is that transfer students come in less well prepared, and therefore do worse than those who began as first-time, full-time freshman, but this was not borne out by the data. Finally, writing was least affected by the intervention, for which there are several possible explanations, and which may be explored in the future. Writing may be the most difficult skill to improve in a short period of time, especially because these students had built up their writing habits over their entire academic careers. It may also be true that by focusing course objectives more heavily on research, writing skills became less important to students.

6.6 Limitations

This study investigated information literacy outcomes in the selected sample of students at one university and although the methods may be instructive, conclusions cannot be applied to other academic institutions in general. The limitation of essay length has been discussed earlier, but deserves mention here. The scorers were trained solely by the researcher, whose perspective and prior experience could shape the emphasis and

direction given to readers. However, this was also a source of consistency that may be necessary to achieve significant reliability.

This study tested only a small portion of the model for evaluating information literacy outcomes in higher education, as discussed in chapter 3. The research model was confirmed to show that prior training and practice frequency were related to performance using direct assessment of constructed responses, yet many aspects of the overall model remain untested.

6.7 A Follow-Up Research Study

The model proposed in chapter 3 indicates the iterative nature of such educational assessments, thus follow-up studies could flow directly from this work. Several areas for productive future research have already been mentioned, but in an effort to put into practice this important element, continuous assessment for improvement, a future study is proposed here. With the success of the diagnostic in improving student learning in this cohort, it is logical to ask if lower-division students would be able to make similar gains using the same information literacy curriculum and assessment program. If similar gains could be achieved in the first year, it would be advantageous to students as a building block for future growth in information literacy and analytical writing skills. In addition, the two most important conclusions of the study imply that changes to the prompt, instruction, and criteria might further improve student outcomes. First, although simply writing the diagnostic at the beginning had only a small positive effect, the assignment produced an unintended consequence. Because the language used in the assignment prompted a majority of students to focus on their opinions rather than on authentic exploration, one might ask if students would do better if the prompt were rewritten

making it clearer that they must explore the question before forming an opinion. Would students' depth of analysis improve if the prompt set up a stronger situation requiring the use of academic research tools for discovery?

Second, the reflective statement was informative, but the language used in the prompt may have resulted in more process description with less introspection. Also, the reflection had not been included in the learning outcomes and evaluation rubric. Would its inclusion, and as a focus of instruction, improve the depth of analysis in student reflections on their work, and thereby their overall information literacy performance? Finally, SAT verbal scores were correlated with diagnostic scores. It is not a common practice among librarians to group students by ability when providing information literacy instruction, but this should be explored further. Would providing "extra or different bibliographic instruction" to students who had weak scores make a significant difference?

6.7.1 Follow-Up Study Research Questions

Some of these questions led to the design of a follow-up study to answer these four questions.

1. **Are the variants of the diagnostic brief essay as effective in assessing and improving higher order information literacy skills in a similar cohort of STEM students? Hypothesis 1: Students taking the same course during a subsequent semester will perform as well or better using a variant of the diagnostic essay.**

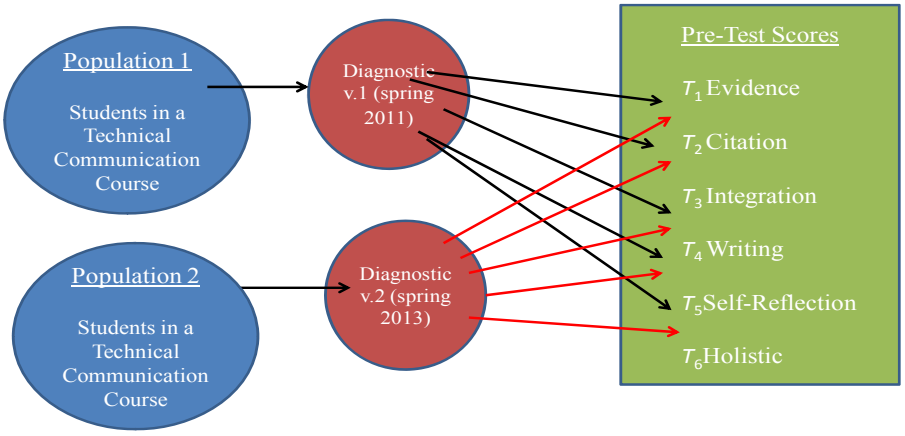


Figure 6-1. Research Model Comparing the Outcomes Using Variant Prompts for the Diagnostic Essay

2. **Will similarly structured assignment and assessment methods be as effective with entering students as with upper-division students? Hypothesis 2: Freshmen composition students taking the diagnostic variant will perform more poorly on average than third-year students, but will make a similar degree of improvement.**

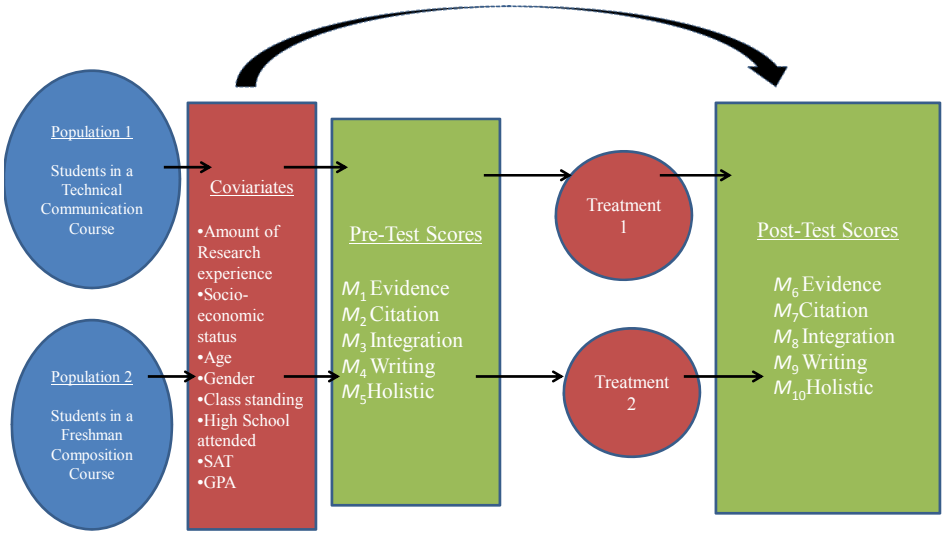


Figure 6-2. Research Model Comparing the Effect of Information Literacy Diagnostic in Freshmen and Upper Division Students

3. **Will the variant of the reflective statement and the addition of the self-awareness criterion have a significant effect on information literacy performance? Hypothesis 3: The revised reflective statement prompt will improve the degree of self-reflection and information literacy scores.**

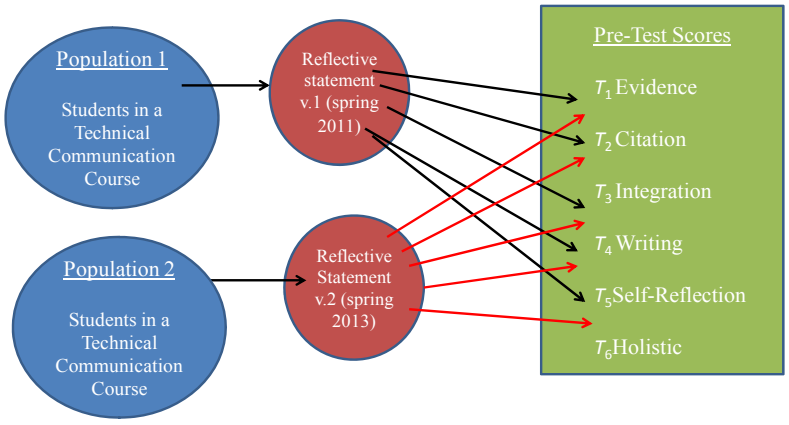


Figure 6-3. Research Model Comparing the Effect of Variant Prompts for the Reflective Statement

4. **What is the effect of having only assignments and no instructional materials?**

   Hypothesis 4: Students will show significant improvement using only assignments as instruction and practice, without a requirement to read or use specific instructional materials.

The research model remains the same as in Figure 4-6; but the new study will compare the performance of students in the same course (using the same instructor and syllabus) with the only difference in treatment being the insertion of the librarian's instructional materials for a random sample of the students. This will control for different course materials, while reexamining the degree to which student use of information literacy instructional materials has any effect.

Librarians and other educators may wish to address the impact of more rigorous assignments, less emphasis on pathfinders, and more emphasis on practice. This

proposed study leaves larger questions to the future. Administrators may wish to consider the overall impact of information literacy outcomes on retention, time to degree, and subsequently professional achievement.

6.8 Closing Statement

An overall model of information literacy assessment in higher education was proposed to isolate important classes of variables affecting learning, including cognitive and affective educational objectives. This study showed that an experimental design and multivariate statistical methods of analysis are feasible. Such methods are necessary if we are to be systematic in our efforts to make sense of the multiple influences of complex variables and sets of variables on information literacy learning. Quantitative assessment of authentic student work product can provide knowledge that enables us to make continuous progress by explaining the relative importance of predictor variables. With this understanding, data driven modifications to the model, and to practice can be made with the goal of continuously increasing the amount of variability explained.

This study showed that information literacy skills are predominantly complex, multipart skills. Application of educational theory and use of an authentic engaging brief research essay and scoring rubric are recommended as a teaching and assessment tool. The information literacy course assignments also served as instruction; that is, by doing the assignment the students were forced to become self-directed learners. Alignment of learning objectives, instructional strategies, and assessment criteria played a major role in this study. Attention to techniques that elicit authentic research should be further studied.

We are entering a period where quantitative data assembled from student work, and student online behavior made available through automated course management

systems, can begin to provide a more detailed understanding of the relationship between instructional design and student learning in information literacy. This study demonstrates that the intuitive understandings that librarians bring to their information literacy work in academic settings can be enhanced by the use of experimental research design, systematic data collection, and multivariate statistical analysis. In addition, information literacy standards and performance indicators are beginning to show their age. They were necessary, but are insufficient to carry the field forward. Investigating relationships between experimental data, the standards, learning theories, and practical instructional design and assessment is the work of the next decade.

APPENDICES

Appendix A. ACRL Standards for Information Literacy (Excerpt)[1]

Information Literacy Defined

Information literacy is a set of abilities requiring individuals to "recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information." Information literacy also is increasingly important in the contemporary environment of rapid technological change and proliferating information resources. Because of the escalating complexity of this environment, individuals are faced with diverse, abundant information choices--in their academic studies, in the workplace, and in their personal lives. Information is available through libraries, community resources, special interest organizations, media, and the Internet--and increasingly, information comes to individuals in unfiltered formats, raising questions about its authenticity, validity, and reliability. In addition, information is available through multiple media, including graphical, aural, and textual, and these pose new challenges for individuals in evaluating and understanding it. The uncertain quality and expanding quantity of information pose large challenges for society. The sheer abundance of information will not in itself create a more informed citizenry without a complementary cluster of abilities necessary to use information effectively.

Information literacy forms the basis for lifelong learning. It is common to all disciplines, to all learning environments, and to all levels of education. It enables learners to master content and extend their investigations, become more self-directed, and assume greater control over their own learning. An information literate individual is able to:

Standard One	Determine the extent of information needed
Standard Two	Access the needed information effectively and efficiently
Standard Three	Evaluate information and its sources critically
Standard Four	Use information effectively to accomplish a specific purpose
Standard Five	Understand the economic, legal, and social issues surrounding the use of information, and access and use information ethically and legally

[1] American Library Association. (2000). *Information literacy competency standards for higher education*. Chicago: Association of College & Research Libraries. Full document is available at http://www.ala.org/acrl/standards/informationliteracycompetency.

Appendix B. Diagnostic Essay and Scoring Rubric

Name:_____ UCID_____

Read This First: Below is a writing task that will provide an idea of your best academic writing and research skills so that you may track your own learning by the end of the semester. An important goal of this class is for you to improve your ability to find and use information, so you can and should refer to outside information in completing this assignment. Use the web and the library as you feel it is appropriate. However, DO NOT consult with others as the purpose is to diagnose your own abilities. Be sure to include your name and UCID in the final submission.

Question statement

As social networking becomes more common, employers have begun to review the pages of current and prospective employees though it may not be a valid measure of work habits or attitude. In the article *No Place to Play: Current Employee Privacy Rights in Social Networking Sites,* Genova (2009) takes the view that employers have a right to monitor employees' information on Facebook. At the same time, many of those seeking employment do not consider this when posting details of their own lives on these types of pages. *Examining Students' Intended Image on Facebook: "What Were They Thinking?!"* an article recently published in the *Journal of Education for Business* shows that people may be disregarding the possible negative consequences of what they post on Facebook. Beyond this, studies have shown that many new employees are spending time on Facebook during the workday without any work-related reason for this access, negatively impacting productivity. Given these issues, do you feel that your activity and information on social networking sites like Facebook should have an impact on your ability to secure a job or maintain your employment?

What To Do

This task has two distinct parts—a formal research essay responding to the question listed above and a reflective statement. You must complete both as Word documents following the submission guidelines for class. Please submit all your work to the assignment tool provided on Moodle labeled "Diagnostic Pretest Submission." Work needs be received by January 24, 2011.

1. Formal Research Essay

For this task you are asked to write a formal persuasive essay that clearly states your position on the given question which is supported with appropriate research and evidence. The essay must be

at least 300 words and have evidence and support provided in the text. You should provide sources cited in the proper APA format within the text and a reference page at the end.

2. Reflective Statement

For this part, you are asked to explain the thinking and process you used to complete the first part of the task. Please explicitly describe how you first engaged with the topic, how you found material to support your points used in the essay, and how you determined these sources to be reliable, trustworthy, and of quality for the essay. This reflective statement can be as long as you feel is needed to fully describe what is asked.

How It Will Be Graded:

When we consider your submission, we will look at the skills needed for writing a task involving location and analysis of information, as well as utilization of the information in argumentation. We will assess the following aspects:

1) Evidence of Research, your ability to access information when needed
2) Evaluation: your ability to evaluate information
3) Citation: your ability to cite your sources correctly
4) Integration: your ability to use the results of your research to make an effective argument
5) Writing: your ability to use appropriate professional written English

SCORING RUBRIC	+ Accomplished – 10 9 8	+ Developing - 7 6 5	+ Poor - 4 3 2	Very Poor 1
Evidence of Research and Quality of Sources	Identifies and selects resources that are relevant and that meet all or nearly all of the important criteria--sensitive to issues of validity, timeliness, and sufficiency. Able to identify valid sources that have been reliably reviewed by those recognized as knowledgeable about the topic at hand, to select sources that offer time-appropriate views on that topic, and to ensure that the sources used are adequate to support the demands of the topic.	Used at least two sources beyond those given. Enough info about the source to indicate it was sought and selected from among several. Quality may be uneven. May have some issues of reliability, sufficiency, relevance, etc. May show some confusion on types of sources.	Alludes to data from at least one source beyond the syllabus, but may miss the key idea completely. Quality of sources may be highly questionable or seriously flawed. May use too few or poor choices.	No evidence that research was conducted. If any sources are used beyond those given there is a superficial or incorrect understanding. May be no evidence that the source was read.
Mechanics of Citation	All cited sources are correct, consistent and in a recognized format (APA/MLA). Citation elements are present, but there may be a minor formatting error.	The citation is missing a minor data element such as author's name or page number, but the source is still findable without undue burden. Appropriate links between in-text and full references.	Citation is missing critical information, such as the journal name or article title. Mostly lacking correct in-text/endnote correspondence. Reference list may not be alphabetized.	Citations are missing, or minimal with only url present, or only the title and author.
Ideas / Integration of Sources into Content	Uses evidence appropriately and effectively, providing sufficient evidence and explanation to convince.	Begins to offer reasons to support its points, perhaps using varied kinds of evidence. Begins to interpret the evidence and explain connections between evidence and main ideas. Its examples bear some relevance.	Often uses generalizations to support its points. May use examples, but they may be obvious or not relevant. Often depends on unsupported opinion or personal experience, or assumes that evidence speaks for itself and needs no application to the point being discussed. Often has lapses in logic.	Depends on clichés or overgeneralizations for support, or offers little evidence of any kind. May be personal narrative rather than essay, or summary rather than analysis.
Writing	Almost entirely free of spelling, punctuation, and grammatical errors. Chooses words for their precise meaning and uses an appropriate level of specificity. Sentence style fits paper's audience and purpose. Sentences are varied, yet clearly structured and carefully focused, not long and rambling. Uses a logical structure appropriate to paper's subject, purpose, audience, thesis, and disciplinary field. Sophisticated transitional sentences often develop one idea from the previous one or identify their logical relations. It guides the reader through the chain of reasoning or progression of ideas.	May contain a few errors, which may annoy the reader but not impede understanding. Generally uses words accurately and effectively, but may sometimes be too general. Sentences generally clear, well structured, and focused, though some may be awkward or ineffective. Indicates a logical progression of ideas and uses fairly sophisticated transitional devices; e.g., may move from least to more important idea. Some logical links may be faulty, but each paragraph clearly relates to paper's central idea.	Usually contains either many mechanical errors or a few important errors that block the reader's understanding and ability to see connections between thoughts. Uses relatively vague and general words, may use some inappropriate language. Sentence structure generally correct, but sentences may be wordy, unfocused, repetitive, or confusing. May list ideas or arrange them randomly rather than using any evident logical structure. May use transitions, but they are likely to be sequential (first, second, third) rather than logic-based. Although each paragraph may relate to central idea, logic is not always clear. Paragraphs have topic sentences but may be overly general, and arrangement of sentences within	Usually contains either many mechanical errors or a few important errors that block the reader's understanding and ability to see connections between thoughts. May contain so many mechanical errors that it is impossible for the reader to follow the thinking from sentence to sentence. May be too vague and abstract, or very personal and specific. Usually contains several awkward or ungrammatical sentences; sentence structure is simple or monotonous. Usually contains many awkward sentences, misuses words, employs

			paragraphs may lack coherence.	inappropriate language.
Holistic & Sophistication	Excels in responding to assignment. Interesting, demonstrates sophistication of thought. Central idea/thesis is clearly communicated, worth developing; limited enough to be manageable. Paper recognizes some complexity of its thesis: may acknowledge its contradictions, qualifications, or limits and follow out their logical implications. Understands and critically evaluates its sources, appropriately limits and defines terms.	A solid paper, responding appropriately to assignment. Clearly states a thesis/central idea, but may have minor lapses in development. Begins to acknowledge the complexity of central idea and the possibility of other points of view. Indicates careful reading of sources, but may not evaluate them critically. Attempts to define terms, not always successfully	Adequate but weaker and less effective, possibly responding less well to assignment. Presents central idea in general terms, often depending on platitudes or clichés. Usually does not acknowledge other views. Indicates basic comprehension of sources, perhaps with lapses in understanding. If it defines terms, often depends on dictionary definitions.	Does not have a clear central idea or does not respond appropriately to the assignment. Thesis may be lacking or too vague or obvious to be developed effectively. Paper may neglect to use any sources or completely misunderstand sources.

Appendix C. Survey and Consent Form

Demographic Information ENG352

2. First name_____
3. Last name _____
4. * UCID (used in your NJIT email address) _____
5. What is the highest educational level attained by either of your parents (for example, your mother OR your father) or the most significant caretaker who influenced you as you were growing up?
    - o 8th grade
    - o high school
    - o some college
    - o graduated college
    - o master's level degree completed
    - o PhD., MD, or other doctoral level degree completed

6. *What is your age? _____

7. Gender: ___ male ___ female
8. What is your current standing at NJIT?
    - o Freshman (0-28 credits)
    - o Sophomore (29-56 credits)
    - o Junior (57-90 credits)
    - o Senior (91+ credits)

9. What is your current GPA? _____
10. What was your High School GPA? _____
11. What was your SAT score? _____
12. Did you transfer to NJIT from another college? ____yes _____ no
13. What is your major? _____
14. We are seeking your permission to make use of the information you provide in this quiz and in this course as part of a research study that will help us improve the quality of teaching at NJIT and possibly elsewhere. No extra work will be required and no identifiable information about you personally or your work will be disclosed. Your participation as part of the research study is voluntary and will not affect your grade in any way.

    You are being invited to participate in a research study about the ability of students to conduct research at the college level. This study is being conducted by me, Davida Scharf, Director of Reference at the Van Houten Library, at the New Jersey Institute of Technology as part of my own doctoral research. You were selected as a possible participant in this study because you are representative of college students at our institution.

    Your participation will not affect your grade in any way.

Your participation will not entail any additional work beyond that normally assigned to all students in order to complete the required coursework.

Your participation is voluntary and the faculty member teaching your course and assigning your grade will not be aware of your decision to participate.

There are no known risks or costs to you if you decide to participate. Only I will be able to identify you or your answers to the survey that follows. Should the research study be published, no identifiable information about you personally will be disclosed.

Your participation is voluntary and your agreement to participate will allow me to use your course assignments and student record for the study. The information collected may not benefit you directly, but may assist NJIT in improving the quality of their instructional programs. Though there are no known risks, neither NJIT nor Rutgers bears any responsibility for any adverse effects that might occur. Study results will be provided to participants upon request.

You are free to decline to answer any particular question in this survey that you do not wish to answer for any reason. You are free to withdraw from the study at any time. You may do so by contacting me for this or for any other reason related to this study. --Davida Scharf, Director of Reference, Van Houten Library, New Jersey Institute of Technology, 323 Martin Luther King Blvd., Newark, NJ 07102, scharf@NJIT.edu.

If you have any questions about your rights as a research subject, you may contact the IRB (a committee that reviews research studies in order to protect research participants) by contacting the IRB Administrator at Rutgers University at: Rutgers University, the State University of New Jersey, Institutional Review Board for the Protection of Human Subjects, Office of Research and Sponsored Programs, 3 Rutgers Plaza, New Brunswick, NJ 08901-8559, Tel: 732-932-0150 ext. 2104, Email: humansubjects@orsp.rutgers.edu

- o I agree
- o I disagree

Appendix D. Instructional Materials

Task 7: Researching, Decision Making, and Evaluating

For this task, you are asked to build on the learning from the previous tasks to begin to apply it to the evaluation of communication packages created by others. You will be examining web pages and are asked to consider questions of quality, reliability, and relevance as you do research and use materials to make decisions and support your work. Start by Reading FUEL Chapter 5 and watch the Creating Effective Packages videos. Then download and read the Librarians expert method for searching on the web. Once all this is done, please read the instructions for the green page quiz and complete it as instructed. Be sure to click the ‘ Submit all and finish ’ button when you are done. This will bring up a message and URL with the librarian’s expert answer for you to read and compare to what you wrote.

Information Literacy Instructional Materials Related To This Assignment (see following pages)

7.1 Librarian's Expert Method for Searching

7.2 Green Web Page Evaluation Instructions and Quiz

7.3 Green Pages Assignment- The Expert's Critiques

- Questions to Ask When Evaluating Sources (text)
- Expert Critique of Lindzen Web Page (pdf)
- Expert Critique of Lindzen Web Page (video)
- Expert Critique of Wikipedia Page on Global Warming (pdf)
- Expert Critique of Lischak Web Page (pdf)
- Expert Critique of EPA Web Page (pdf)

7.1 Searching and Evaluating Websites – *Green* Example

What you will review here is the start of information literacy and good research skills as well as a means to improve your ability to recognize and critique better from worse. I want to stress that the material found on the Internet or any source can vary in quality and usefulness depending on the situation, how you will use it, and your ability to judge its quality for the intended use. Our aim for this first assignment is to start your investigation into what you do now and how it can be improved by knowing what to look for and what questions to ask. It will also hopefully show you more sources of good information and how you can search more effectively for quality work that will inform and support all your projects. Please read through each part of this project and do them in order. If you skip ahead, you can finish but may not benefit from the learning objects provided as intended and so have more difficulty later in the class.

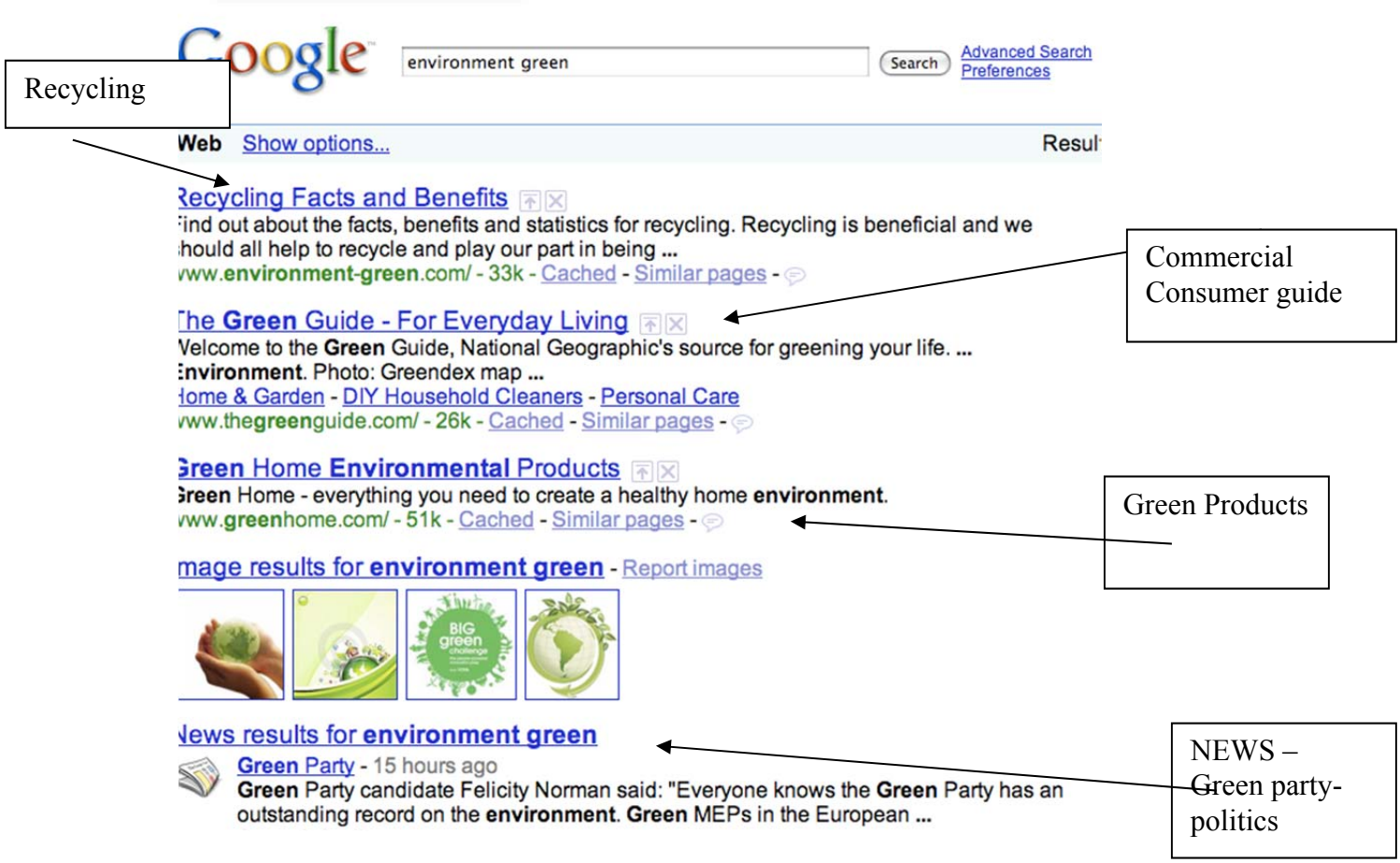
This file contains a description of the thinking and actions that an expert follows as she uses search engines to find support for an assignment. Imagine that I assigned an essay about the 'green' debate and revision of a topic to edit on Wikipedia. As part of that assignment, each student would have to find sources related to something 'green' in order to complete the work.

I have asked Davida Scharf, a NJIT reference Librarian to describe her thinking as she engages with the use of Google so that you might compare this with what you do as you initiate a search. We will then ask you to write out your opinions and ideas as you move forward. This will allow you to compare your work with what an expert in research and information literacy does in order to compare your methods and critical thinking.

WEBSITES TO CONSIDER AND HOW I GOT THERE

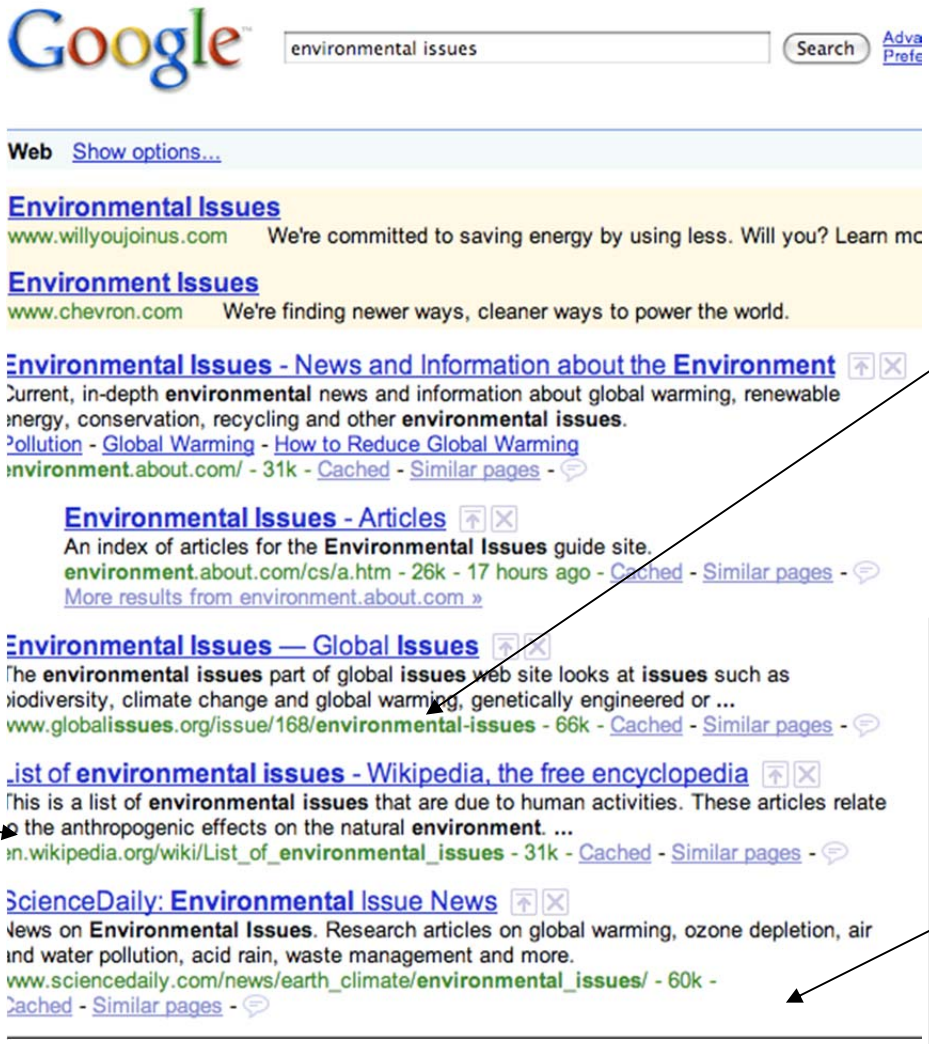
Prepared by Davida Scharf, Librarian
New Jersey Institute of Technology

First I searched Google using the words ' environment green ' The results weren't that helpful. The top few results were about recycling and about green living, also politics from the news result, rather than global environmental issues or technology—which is what I had in mind.



I realized the word green was bringing up consumer type sites so I redid my search using the terms ' environmental issues' This resulted in improved results and some ideas for narrowing the topic using additional keywords.

Dr Lipuma's Note: By reading the summary, looking at the URL, and clicking on some pages to read more that look promising, the expert gathers helpful information. By knowing what to look for and having a more advanced sense of what may be of quality, the expert can be more effective at searching. However, when beginning the search, the expert builds knowledge about the topic and the terminology that is associated with it to both learn about the topic as well as improve the ability to search it and find sources of quality information.

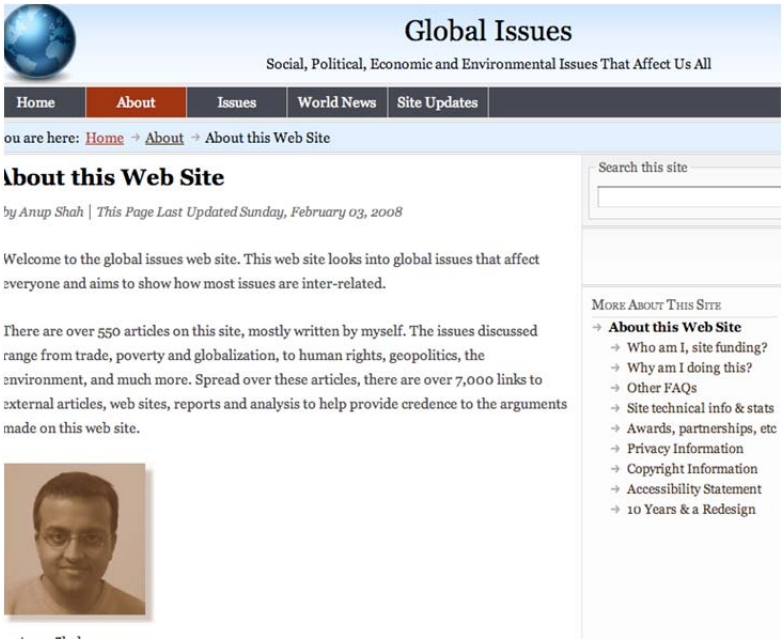


About.com and Wikipedia are good places to get started… the basics, scope of topic and subtopics, get new ideas, additional keywords, links to reliable sources…

International organization??

Global warming, ozone depletion, pollution, acid rain, waste management, etc. This gave me the idea to use a clustering search engine… clusty.com…see below

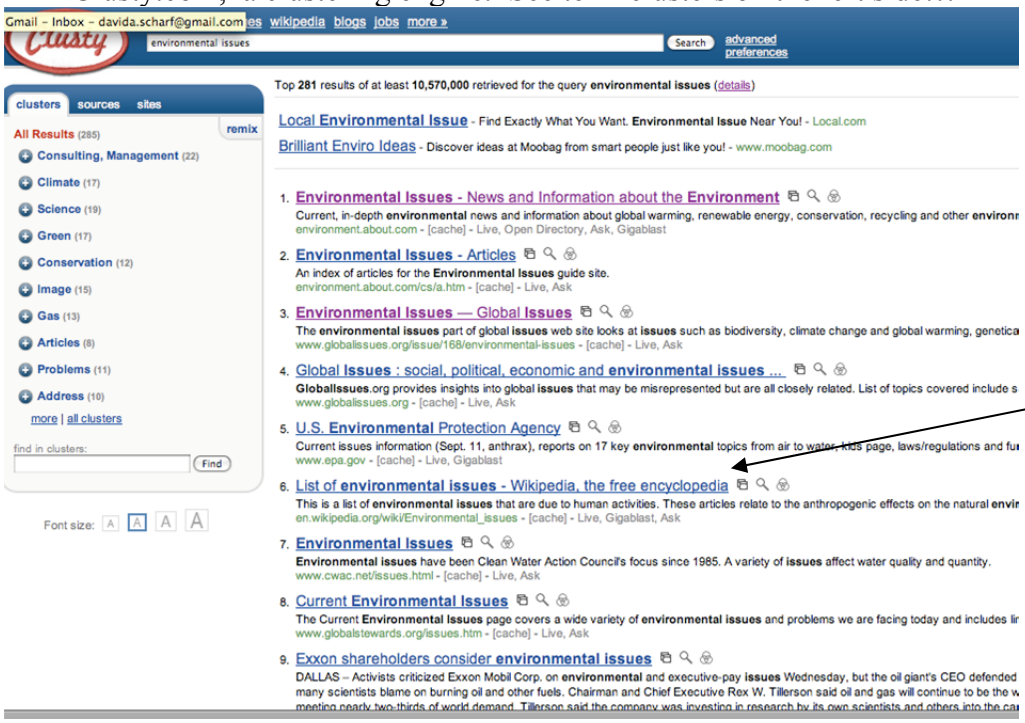
http://www.globalissues.org/ looked like it might be some kind of international organization. Look what it turned out to be! What is useful about it? Do you have any reservations about the website? Why?



Dr Lipuma's Note: The expert does not necessarily stay within one search engine, or source. By gathering information from different sources, the expert is able to understand how others discuss the topic and so see how it might be more easily searched to find what is being sought.

Since Google provided me with some terms to search, I used another search engine

TRY Clusty.com, a clustering engine. See term clusters on the left side…

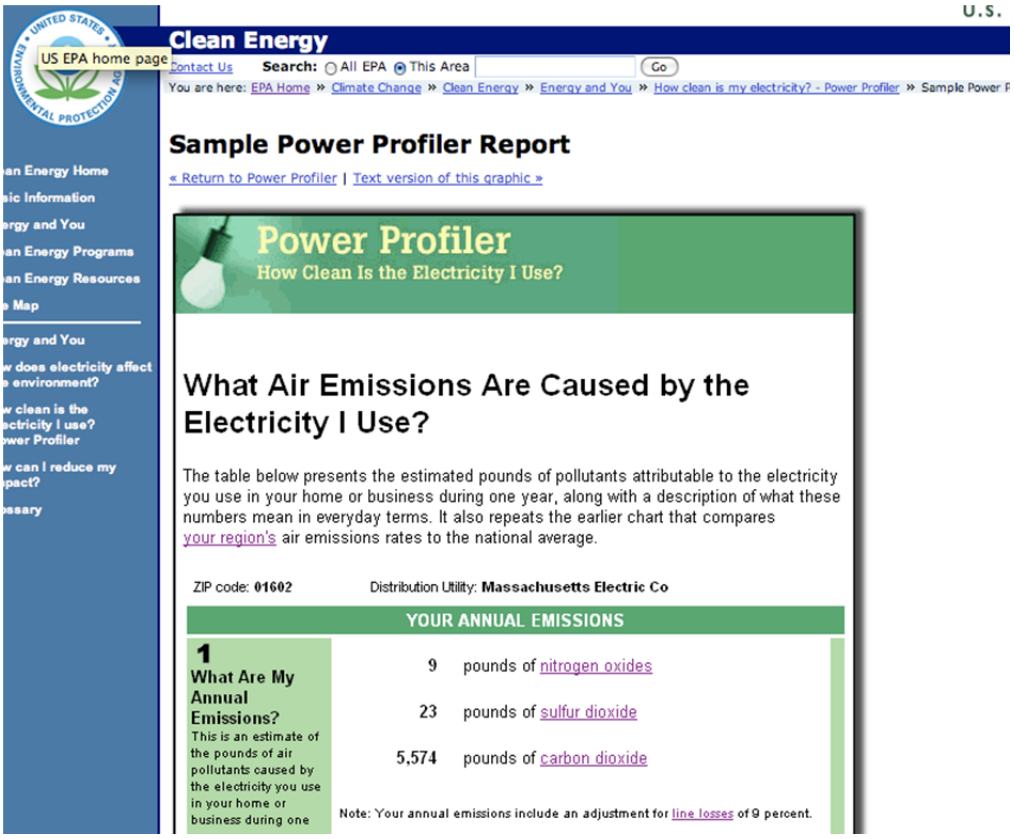


EPA.gov

(Govt sources at USA.gov.)

Dr Lipuma's Note: Even though a search engine returns a result that might look promising, and the source is of quality, it may not be useful for your purposes or refined enough if used without further searching. Search engines do not know what you want and can only return results based upon your ability to ask the right question within the parameters of the programming and algorithms governing the search protocol. Popular web pages or pages that have paid the search engine may move to the top of the list of returned results but might not be what you are looking for. In other cases, pages that have many links to useful information may be found but if you do not dig deeper, the page itself will not be useful except as a portal.

EPA.gov is a portal. Can you find a page on emissions or on clean coal here?
http://www.epa.gov/cleanenergy/energy-and-you/report-large.html



Questions of Quality

How do you judge the quality of what you have found? Since so much can be posted on the internet without any verification, it is important to consider various aspects of the quality of the source and the content of what you find. To assist with this, the library has provided the following page with areas to consider and questions to ask about each of these areas.
http://library.njit.edu/researchhelpdesk/howto/evaluate.php

Your view of quality of internet Pages

One reason to use sources that have been reviewed or verified is to assist you to be confident in the quality of the source and the content. However, this is not always a guarantee. Journals, books, databases and many other sources are a good way to have more confidence in what is presented, but you must still ask questions about sources. If the material is posted on the web, the need to verify quality is even more important. In other cases, the source is very reliable and of high quality but tee method of presenting the material makes it unclear or difficult to access. It is these two ideas that will be explored further throughout the course.

Please look at the following pages to compare how you might judge the quality of the source. Post your answers in the quiz on Moodle and you will receive a link to the page where the librarian provided her opinions about each of these pages.

Global Warming: ***Yale Center for the Study of Globalization***
http://www.independent.org/publications/article.asp?id=1714

Global Warming: Church of God
http://www.realtruth.org/articles/443-gwrfa.html

Global Warming EPA.gov
http://www.epa.gov/climatechange/science/stateofknowledge.html

7.2 Green Web Page Evaluation Instructions / Quiz

Now that you have read the librarian's method of searching for a green topic please find the following pages and review what they say as if you found them as support for a paper on Global Warming. You need to determine what you think of the quality and reliability of the content on each page and the quality of each source.

Global Warming: Yale Center for the Study of Globalization
http://www.independent.org/publications/article.asp?id=1714

Global Warming: Church of God
http://www.realtruth.org/articles/443-gwrfa.html

Global Warming EPA.gov
http://www.epa.gov/climatechange/science/stateofknowledge.html

Once you have reviewed these pages, please complete the Green Page Reflection Quiz. For this quiz, you just need to comment on what you feel the quality of each of the three web pages are and why you feel this way. When you complete the quiz you will be given the URL to the comments about each page given by the Librarian our expert on reviewing sources. You should read what the librarian said to compare the expert view of the quality and method for determining this to your own interpretation.

7.3 Green Pages Assignment- The Expert's Critiques

full text of materials below available at
http://researchguides.njit.edu/content.php?pid=285943&sid=2499704)

- Questions to Ask When Evaluating Sources (text)
- Expert Critique of Lindzen Web Page (pdf)
- Expert Critique of Lindzen Web Page (video)
- Expert Critique of Wikipedia Page on Global Warming (pdf)
- Expert Critique of Lischak Web Page (pdf)
- Expert Critique of EPA Web Page (pdf)

Task 8: Finding and Evaluating Sources

Task 8 Instructions

8.1 Watch the Decision Making Videos Below:
Decision Making
Data, Information, Knowledge
Dealing with Uncertainty

8.2 Watch the Videos Below on Finding Good Sources
Demonstration of web searching for reliable results (3:54) file
Links to Tips on Advanced Google Searching-from a Princeton Librarian
Finding the Full Text of an Article (2:26)
Finding E-Books (2:00) Powerpoint presentation
Understanding Citations (7:19)

8.3 Watch these videos about Wikipedia
WISER - Introduction to Wikipedia Resource
WISER - What Makes a Good Page Resource
External Links vs. References in Wikipedia Resource

8.4 Summary, Annotation, and Evaluation

8.5 Wikipedia Source Evaluation Explanation
The Wikipedia Article on Coaxial Cable (pdf)
Gross-Sample Article Assessment by Librarian (doc)

8.6 Wikipedia Source Evaluation Submission (assignment)

8.7 Wikipedia Source Evaluation Expert Answer (doc)

Task 10: WISER Project Proposal

Task 10 Instructions Resource
10.1-Wikipedia Request for Proposals (RFP) Word document
10-2-Topic Selection Forum
10.3- Librarian's Recommended Resources for the WISER Project
WISER Introduction Resource
WISER - Good Page Example Resource
Librarian's Wikipedia Intro in PPT file Powerpoint presentation
How the Librarian chose a topic Powerpoint presentation
Parts of a Citation Word document
APA Style manual for Citations PDF document
10.4-Submit your WISER Proposal Assignment
Check WISER Proposal for Plagiarism- Assignment
10.4b-WISER Proposal Regrading Submissions Assignment
WISER Proposal--Essay Sample #1 Word document
WISER Proposal--Essay Sample #2 Word document
WISER Proposal--Letter Sample Word document
WISER Proposal--Memo Sample Word document
10.5-Information Literacy Rubric Word document

Task 11: WISER Project Peer Review

Task 11 Instructions Resource
11.1-Download the Peer Review Explanation Word File Word document
11.2-WISER Peer Review Report Forum
Live Chat Room Wimba Classroom
11.3-Final Memo Reporting on WISER Project Assignment

Task 13: Persuasive Research Essay

Task 13 Instructions Resource
13.1-Persuasive Research Essay Instructions Word document
13.2-Persuasive Research Essay submission Assignment
13.3 Research Essay Plagiarism Check Assignment

Task 14: Final Showcase Submissions

Task 14 Instructions Resource
14.1-Final Showcase Narrative Template Spring 2011 Word document
14.1b-Overview Word document
Please name the file correctly with YOUR last name before submitting.
14.2-Final Showcase Submissions Assignment

Appendix E. Scoring Criteria Mapped to ACRL Standards and Bloom's Taxonomy

Local Criteria	Evidence of Local Criteria	*ACRL Standards*	Applicable ACRL Performance Indicators	ACRL-Krathwohl
Evidence of Research and Quality of Sources	Identifies and selects resources that are relevant and that meet all or nearly all of the important criteria--sensitive to issues of validity, timeliness, and sufficiency. Able to identify valid sources that have been reliably reviewed by those recognized as knowledgeable about the topic at hand, to select sources that offer time-appropriate views on that topic, and to ensure that the sources used are adequate to support the demands of the topic.	*1. Determines the nature and extent of the information needed.* *2. Accesses needed information effectively and efficiently.* *3. Evaluates information and its sources critically and incorporates selected information into his or her knowledge base and value system.*	1.1 Defines the need for information 1.2 Identifies a variety of types of formats of potential sources of information 1.3 Considers the costs and benefits of acquiring the needed information 1.4 Re-evaluates the nature and extent of the information need (Improvement in post-test only) 2.1 Selects the most appropriate investigative methods or information retrieval systems for accessing the needed information 2.2 Constructs and implements effectively-designed search strategies 2.3 Retrieves information online or in person using a variety of methods 2.4. Refines the search strategy if necessary 3.2. Articulates and applies initial criteria for evaluating both the information and its sources	1.1 Analyze, Create 1.2 Apply 1.3 Analyze 1.4 Evaluate 2.1 Evaluate 2.2 Create 2.3 Apply 2.4 Evaluate 3.2 Analyze
Mechanics of Citation	All cited sources are correct, consistent and in a recognized format (APA/MLA). Citation elements are present, but there may be a minor formatting error.	*2. Accesses needed information effectively and efficiently.* *5. Understands many of the economic, legal, and social issues surrounding the use of information and accesses and uses information ethically and legally.*	2.3. Retrieves information online or in person using a variety of methods. 2.5. Extracts, records, and manages the information and its sources 5.2. Follows laws, regulations, institutional policies, and etiquette related to the access and use of information 5.3. Acknowledges the use of information sources in communicating the product or performance	2.3 Apply 2.5 Apply 5.2 Apply 5.3 Apply

Local Criteria	Evidence of Local Criteria	*ACRL Standards*	Applicable ACRL Performance Indicators	ACRL-Krathwohl
Ideas / Integration of Sources into Content	Uses evidence appropriately and effectively, providing sufficient evidence and explanation to convince.	*3. Evaluates information and its sources critically and incorporates selected information into his or her knowledge base and value system.* *4. Individually or as a member of a group, uses information effectively to accomplish a specific purpose.* *5. Understands many of the economic, legal, and social issues surrounding the use of information and accesses and uses information ethically and legally.*	3.1. Summarizes the main ideas to be extracted from the information gathered 3.3. Synthesizes main ideas to construct new concepts 3.4. Compares new knowledge with prior knowledge to determine the value added, contradictions, or other unique characteristics of the information. 4.1. Applies new and prior information to the planning and creation of a particular product or performance 5.1. Understands many of the ethical, legal and socio-economic issues surrounding information and information technology	3.1 Analyze 3.3 Create 3.4 Evaluate 4.1 Create 5.1 Understand
Writing	Almost entirely free of spelling, punctuation, and grammatical errors. Chooses words for their precise meaning and uses an appropriate level of specificity. Sentence style fits paper's audience and purpose. Sentences are varied, yet clearly structured and carefully focused, not long and rambling. Uses a logical structure appropriate to paper's subject, purpose, audience, thesis, and disciplinary field. Sophisticated transitional sentences often develop one idea from the previous one or identify their logical relations. It guides the reader through the chain of reasoning or progression of ideas.	*3. Evaluates information and its sources critically and incorporates selected information into his or her knowledge base and value system.* *4. Individually or as a member of a group, uses information effectively to accomplish a specific purpose.*	3.1. Summarizes the main ideas to be extracted from the information gathered 3.3. Synthesizes main ideas to construct new concepts 3.4. Compares new knowledge with prior knowledge to determine the value added, contradictions, or other unique characteristics of the information. 4.1. Applies new and prior information to the planning and creation of a particular product or performance 4.3. Communicates the product or performance effectively to others	3.1 Analyze 3.3 Create 3.4 Evaluate 4.1 Create 4.3 Create

Local Criteria	Evidence of Local Criteria	*ACRL Standards*	Applicable ACRL Performance Indicators	ACRL-Krathwohl
Holistic & Sophistication	Excels in responding to assignment. Interesting, demonstrates sophistication of thought. Central idea/thesis is clearly communicated, worth developing; limited enough to be manageable. Paper recognizes some complexity of its thesis: may acknowledge its contradictions, qualifications, or limits and follow out their logical implications. Understands and critically evaluates its sources, appropriately limits and defines terms.	*3. Evaluates information and its sources critically and incorporates selected information into his or her knowledge base and value system.* *5. Understands many of the economic, legal, and social issues surrounding the use of information and accesses and uses information ethically and legally.*	3.5. Determines whether the new knowledge has an impact on the individual's value system and takes steps to reconcile differences 3.6. Validates understanding and interpretation of the information through discourse with other individuals, subject-area experts, and/or practitioners (peer-review) 3.7. Determines whether the initial query should be revised (Improvement in post-test only) 5.1. Understands many of the ethical, legal and socio-economic issues surrounding information and information technology	3.5 Evaluate 3.6 Create 3.7 Evaluate 5.1 Understand

Appendix F: Institutional Research Board Approvals

RUTGERS UNIVERSITY
Office of Research and Sponsored Programs
ASB III, 3 Rutgers Plaza, Cook Campus
New Brunswick, NJ 08901

September 10, 2010

P.I. Name: Scharf
Protocol #: E11-053

Davida Scharf
330 Highwood Ave.
Leonia NJ 07605

Dear Davida Scharf:

Notice of Exemption from IRB Review

Protocol Title: "Information Literacy"

The project identified above has been approved for exemption under one of the six categories noted in 45 CFR 46, and as noted below:

Exemption Date: 8/25/2010 **Exempt Category:** 1

This exemption is based on the following assumptions:

- **This Approval** - The research will be conducted according to the most recent version of the protocol that was submitted.
- **Reporting** – ORSP must be immediately informed of any injuries to subjects that occur and/or problems that arise, in the course of your research;
- **Modifications** – Any proposed changes MUST be submitted to the IRB as an amendment for review and approval prior to implementation;
- **Consent Form (s)** – Each person who signs a consent document will be given a copy of that document, if you are using such documents in your research. The Principal Investigator must retain all signed documents for at least three years after the conclusion of the research;

Additional Notes: **None**

Failure to comply with these conditions will result in withdrawal of this approval.

The Federalwide Assurance (FWA) number for Rutgers University IRB is FWA00003913; this number may be requested on funding applications or by collaborators.

Sincerely yours,

Sheryl Goldberg
Director of Office of Research and Sponsored Programs
egraser@grants.rutgers.edu

cc: Daniel O'Conner

NJIT New Jersey's Science & Technology University

Institutional Review Board: HHS FWA 00003246
Notice of Approval
IRB Protocol Number: E148-11

Principal Investigators: Davida Scharf
Library

Title: Information Literacy
Performance Site(s): NJIT:

Type of Review: FULL [ ] EXPEDITED [x]

Type of Approval: NEW [x] RENEWAL [ ] REVISION [ ]

Approval Date: February 14, 2011 Expiration Date: February 13, 2012

1. **ADVERSE EVENTS:** Any adverse event(s) or unexpected event(s) that occur in conjunction with this study must be reported to the IRB Office immediately (973) 596-5825.

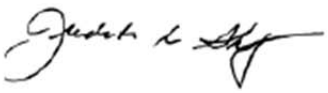
2. **RENEWAL:** Approval is valid until the expiration date on the protocol. You are required to apply to the IRB for a renewal prior to your expiration date for as long as the study is active. It is your responsibility to ensure that you submit the renewal in a timely manner.

3. **CONSENT:** All subjects must receive a copy of the consent form as submitted. Copies of signed consent forms must be kept on file with the principal investigator.

4. **SUBJECTS:** Number of subjects approved: 1,000

5. The investigator(s) did not participate in the review, discussion, or vote of this protocol.

6. **APPROVAL IS GRANTED ON THE CONDITION THAT ANY DEVIATION FROM THE PROTOCOL WILL BE SUBMITTED, IN WRITING, TO THE IRB FOR SEPARATE REVIEW AND APPROVAL.**



Judith Sheft, IRB Chair,

REFERENCES

ACT. Collegiate Assessment of Academic Proficiency (CAAP). (2012) Retrieved May 26, 2012, from http://www.act.org/caap/.

Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine, 68*(1), 52–81.

Albitz, R. (2007). The what and who of information literacy and critical thinking in higher education. *Portal: Libraries & the Academy, 7*(1), 97–110.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association

American Library Association (ALA). (1989). *Presidential Committee on Information Literacy: Final Report.* Chicago: Author. Retrieved from http://www.ala.org/ala/mgrps/divs/acrl/publications/whitepapers/presidential.cfm.

American Library Association. (2000). *Information literacy competency standards for higher education*. Chicago: Association of College & Research Libraries.

American Library Association. (2004). Press Release: ACRL information literacy standards endorsed by the Council of Independent Colleges. Retrieved May 27, 2012, from http://www.ala.org/Template.cfm?Section=archive&template=/contentmanagement/contentdisplay.cfm&ContentID=59876.

Anderka, M., & Stein, B. (2012). A breakdown of quality flaws in Wikipedia. Paper presented at the Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality, Lyon, France.

Andretta, S. (2011). Information literacy: A term whose time has passed? *Journal of Information Literacy, 5*(1), 1–4.

Arp, L. (1990). Information literacy or bibliographic instruction semantics or philosophy? *RQ: Research Quarterly, 30*(1), 46–49.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago: University of Chicago Press.

Association of College & Research Libraries. PRIMO; Peer-reviewed instructional materials online. Retrieved April 20, 2012, from http://www.ala.org/acrl/aboutacrl/directoryofleadership/sections/is/iswebsite/projpubs/primo.

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel, 25*(4), 297–308.

Astin, A. W. (1993). Forging the ties that bind: The dilemma of the modern university. *College Board Review,* (165), 1–15, 26–27.

Badke, W. B. (2011). *Research strategies: Finding your way through the information fog*. Bloomington, IN: iUniverse.

Bailin, A., & Grafstein, A. (2005). The evolution of academic libraries: The networked environment. *Journal of Academic Librarianship, 31*(4), 317–323.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.

Banta, T. (2008). Trying to clothe the emperor. *Assessment Update, 20*(2), 3–4.

Bar, T., & Zussman, A. (2012). Partisan grading. *American Economic Journal: Applied Economics*, 4(1), 30-48.

Barkey, P. (1965). Patterns of student use of a college library. *College & Research Libraries, 26*(2), 115–118.

Bazerman, C., & Russell, D. R. (1994). *Landmark essays on writing across the curriculum*. Davis, CA: Hermagoras Press.

Bean, J. P., & Eaton, S. B. (2000). A psychological model of college student retention. In J. M. Braxton (Ed.), *Reworking the student departure puzzle: New theory and research on college student retention* (pp. 73–89). Nashville, TN: Vanderbilt University Press.

Beile-O'Neil, P. M. B. (2005). *Development and validation of the Beile test of information literacy for education (B-TILED).* (Ph.D. diss. University of Central Florida, Orlando). Retrieved December 31, 2013, from http://pennybeile.net/uploads/B-tiled_diss.pdf.

Beile, P., Dziuban, C., Katz, I., & Salem, J. (2010). iSkills/SAILS correlation study. Paper presented at the American Library Association, Washington, DC. Retrieved from http://eprints.rclis.org/bitstream/10760/15859/6/ebss_poster_session.pdf.

Benjamin, R., Chun, M., Hardison, C., Hong, E., Jackson, C., Kugelmass, H., . . . Shavelson, R. (2009). Returning to learning in an age of assessment: Introducing the rationale of the Collegiate Learning Assessment (full report). New York: Council for Aid to Education.

Berger, P., & Luckmann, T. (1991). *The social construction of reality: A treatise in the sociology of knowledge*. London: Penguin Books.

Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher Education Research & Development, 18*(1), 57–75. doi: 10.1080/0729436990180105.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (4th ed.). New York: Open University Press/McGraw-Hill.

Bloom, B. S. (1968). Learning for mastery. Instruction and curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation Comment, 1*, 12.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Bloom, B. S., Krathwohl, D. R., & Masia, B. B. (1956). *Taxonomy of educational objectives : The classification of educational goals*. New York: D. McKay.

Bonanno, H. (2002). *Standing the test of time: revisiting a first year diagnostic procedure.* Paper presented at the Sixth Pacific Rim Conference on First Year in Higher Education: Changing agendas "Te ao hurihuri", Brisbane, Australia.

Bonanno, H., & Jones, J. (2007). *The MASUS procedure: Measuring the academic skills of university students: A diagnostic assessment* (rev. ed.). Sydney, Australia: Learning Centre, University of Sydney.

Bonwell, C. C., & Eison, J. A. (1991). Active learning: Creating excitement in the classroom. 1991 ASHE-ERIC Higher Education Report No. 1. Washington, DC: The George Washington University, School of Education and Development.

Booth, W. C., Colomb, G. G., & Williams, J. M. (2011). *The craft of research*. Chicago: University of Chicago Press.

Bostick, S. L. (1992). *The development and validation of the Library Anxiety Scale.* Ph.D. diss., Wayne State University. Retrieved from http://digitalcommons.wayne.edu/dissertations/AAI9310624 (Paper AAI9310624).

Bowles-Terry, M., Davis, E., & Holliday, W. (2010). Writing information literacy revisited: Application of theory to practice in the classroom. *Reference & User Services Quarterly, 49*(3), 225–230.

Boyer, E. L. (1987). *College: The undergraduate experience in America*. New York: Harper & Row.

Branscomb. (1940). *Teaching with books: A study of college libraries*. Chicago: Association of American Colleges.

Brasley, S. S. (2008). Effective librarian and discipline faculty collaboration models for integrating information literacy into the fabric of an academic institution. *New Directions for Teaching and Learning, 2008*(114), 71–88. doi: 10.1002/tl.318

Braxton, J., Milem, J., & Sullivan, A. (2000). The influence of active learning on the college student departure process: Toward a revision of Tinto's theory. *Journal of Higher Education, 71*(5), 569–590.

Breivik, P. S. (1974). *Effects of library-based instruction in the academic success of disadvantaged college freshmen.* Ph.D., Columbia University, New York. Retrieved from http://clio.cul.columbia.edu:7018/vwebv/holdingsInfo?bibId=2363617.

Bridgeman, B., Trapani, C., & Bivens-Tatum, J. (2011). Comparability of essay question variants. *Assessing Writing, 16*(4), 237–255. doi: 10.1016/j.asw.2011.06.002.

Broughton, J. (2008). *Wikipedia: The missing manual*. Sebastopol: O'Reilly Media. http://en.wikipedia.org/wiki/Help:Wikipedia:_The_Missing_Manual

Bruce, C. (1997). *The seven faces of information literacy*. Adelaide, Australia: Auslib Press

Bruce, C. (2004). Information literacy as a catalyst for educational change: A background paper. (reprint of Bruce's paper in the proceedings of the UNESCO Information Literacy Meeting of Experts). Paper presented at the Lifelong Learning: Whose responsibility and what is your contribution?", the 3rd International Lifelong Learning Conference, Yeppoon, Queensland, Australia. Retrieved from http://eprints.qut.edu.au.

Bruner, J. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Bundy, A., ed. (2004). *Australian and New Zealand information literacy framework: principles, standards and practice.* (2nd ed.). Adelaide: Australian and New Zealand Institute for Information Literacy (ANZIIL), Council of Australian University Librarians (CAUL).

Burkhardt, J. M. (2007). Assessing library skills: A first step to information literacy. *Portal: Libraries & the Academy, 7*(1), 25–49.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The criterion online writing service. *AI Magazine, 25*(3), 27.

Buschman, J. (2009). Information literacy, "new" literacies, and literacy. *Library Quarterly, 79*(1), 95–118.

Callister, P. D. (2010). Time to blossom: An inquiry into Bloom's Taxonomy as a means to ordered legal research skills. *Law Library Journal, 102*(2), 191–218.

Cameron, L., Wise, S. L., & Lottridge, S. M. (2007). The development and validation of the information literacy test. *College & Research Libraries, 68*(3), 229–236.

Chesney, T. (2006, November 6). An empirical examination of Wikipedia's credibility. *First Monday*, 11(11),

Cheung, C. K. (2002). Assessing university students' general and specific critical thinking. *College Student Journal, 36*(4), 504.

Choinski, E., & Emanuel, M. (2006). The one-minute paper and the one-hour class: Outcomes assessment for one-shot library instruction. *Reference Services Review, 34*(1), 148-155.

Choinski, E., Mark, A. E., & Murphey, M. (2003). Assessment with rubrics: An efficient and objective means of assessing student outcomes in an information resources class. *Portal: Libraries & the Academy, 3*(4), 563–575.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155.

Colbert, S. (July 31, 2006). The Word—Wikiality: On Wikipedia, we can create a reality that we can all agree on—the reality we just agreed on Retrieved June 27, 2012, from http://www.colbertnation.com/the-colbert-report-videos/72347/july-31-2006/the-word---wikiality.

Colvin, J., & Keene, J. (2004). Supporting undergraduate learning through the collaborative promotion of e-journals by library and academic departments. *Information Research, 9*(2). Retrieved December 31, 2013 from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1413/1331

Cook, C., & Maciel, M. (2010). A decade of assessment at a research-extensive university library using LibQUAL+. *Research Library Issues, 271*(August), 4–12.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30*(2), 116–127.

Cummings, R. E. (2009). *Lazy virtues: Teaching writing in the age of Wikipedia*. Nashville, TN: Vanderbilt University Press.

D'Angelo, B. J., & Maid, B. M. (2004). Moving beyond definitions: Implementing information literacy across the curriculum. *Journal of Academic Librarianship, 30*(3), 212–217.

Daniels, E. (2010). Using a targeted Rubric to deepen direct assessment of college students' abilities to evaluate the credibility of sources. *College & Undergraduate Libraries, 17*(1), 31–43.

Deess, Perry. (2010, November 17). *Entering Student Surveys: Fall 2010.* Powerpoint. Newark: NJ: New Jersey Institute of Technology. Institutional Research and Planning. Retrieved May 29, 2012, from http://www.njit.edu/irp/reports/2010/Entering_Students_Grad___Undergerad_2010_v2_shortened_final.ppt

Dervin, B., & Nilan, M. (1986). Information needs and uses. *Annual Review of Information Science & Technology, 21*, 3–33.

Detlor, B., Julien, H., Willson, R., Serenko, A., & Lavallee, M. (2011). Learning outcomes of information literacy instruction at business schools. *Journal of the American Society for Information Science & Technology, 62*(3), 572–585. doi: 10.1002/asi.21474.

Dick, W., Carey, L., & Carey, J. O. (2008). *The systematic design of instruction* (7th ed.). New York: Allyn & Bacon.

Diederich, P. B. (1974). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.

Diller, K. R., & Phelps, S. F. (2008). Learning outcomes, portfolios, and rubrics, oh my! authentic assessment of an information literacy program. *Portal: Libraries & the Academy, 8*(1), 75–89.

Durst, R. K. (2006). Writing at the postsecondary level. In P. Smagorinsky (Ed.), *Research on composition: Multiple perspectives on two decades of change* (pp. 78–107). New York: Teachers College Press.

Elliot, N. (2005). *On a scale: A social history of writing assessment in America*. New York: Peter Lang.

Elliot, N., Briller, V., & Joshi, K. (2007). *Analytic portfolio assessment: A program development model. Journal of Writing Assessment* (3)1: 5-29.

Elliot, N., Briller, V., & Joshi, K. (2007). Portfolio assessment: Quantification and community. *Journal of Writing Assessment, 3*(1), 5–30.

Elliot, N., Kilduff, M., & Lynch, R. (1994). The assessment of technical writing. *Journal of Technical Writing & Communication 24*(1), 19–36.

Elmborg, J., & Hook, S. (2005). *Centers for learning: Writing centers and libraries in collaboration*. Chicago: Association of College & Research Libraries.

Elmborg, J. K. (2003). Information literacy and writing across the curriculum: Sharing the vision. *Reference Services Review, 31*(1), 68–80.

Elmborg, J. K., & Hook, S. (2005). *Centers for learning: Writing centers and libraries in collaboration*. Chicago: Association of College & Research Libraries.

Emig, J. (1971). *The composing processes of twelfth graders*. Urbana, IL: National Council of Teachers of English.

Emmons, M., & Martin, W. (2002). Engaging conversation: Evaluating the contribution of libary instruction to the quality of student research. *College & Research Libraries, 63*(6), 545–560.

Emmons, M., Martin, W., Botts, C., & Amundson, C. (2010a). Engaging sources: Information literacy and the freshman research paper (Part I). *LOEX Quarterly, 36*(4), 5.

Emmons, M., Martin, W., Botts, C., & Amundson, C. (2010b). Engaging Sources: information literacy and the freshman research paper (Part II). *LOEX Quarterly, 37*(2), 4.

Emmons, M., & Wilkinson, F. C. (2011). The academic library impact on student persistence. *College & Research Libraries, 72*, 128–149.

Erling, E. J., & Richardson, J. T. E. (2010). Measuring the academic skills of university students: Evaluation of a diagnostic procedure. *Assessing Writing, 15*(3), 177–193. doi: 10.1016/j.asw.2010.08.002.

Farber, E. I. (1974). Library instruction throughout the curriculum: Earlham College program. In J. Lubans (Ed.), *Educating the library user* (pp. 145-162). New York: Bowker.

Farmer, L. J. (2007). Developmental and social-emotional behavior and information literacy. In D. Nahl & D. Bilal (Eds.), *Information and emotion: The emergent affective paradigm in information behavior research and theory* (pp. 99–119). Medford, NJ: Information Today.

Feldman, K. A. (1976). The superior college teacher from the students' view. *Research in Higher Education, 5*(3), 243–288.

Feldman, K. A., & Newcomb, T. (1969). *The impact of college on students*. San Francisco, CA: Jossey-Bass.

Fisher, K. E., Erdelez, S., & Mckechnie, L. (2005). *Theories of information behavior*. Medford, NJ: Information Today.

Fister, B. (1993). Teaching the rhetorical dimensions of research. *Research Strategies, 11*(4), 211–219.

Flower, L. S., & Hayes, J. R. (1979). *A process model of composition: Document Design Project*. Technical Report No. 1. Washington, DC.: National Institute of Education, Dept. of Health Education and Welfare.

Forte, A., & Bruckman, A. (2006). *From Wikipedia to the classroom: Exploring online publication and learning*. Proceedings from ICLS ’06 the 7th International Conference on Learning Sciences (pp. 182-188), Bloomington, IN: International Society of the Learning Sciences.

Forte, A., & Bruckman, A. (2006, June 27-July 1). *From Wikipedia to the classroom: exploring online publication and learning.* Paper presented at the 7th International Conference on Learning Sciences, Bloomington, IN.

Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012). Defining authentic classroom assessment. *Practical Assessment, Research & Evaluation, 17*(2), 2.

George, M. W. (2008). *The elements of library research: What every student needs to know*. Princeton, NJ: Princeton University Press.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature, 438*(7070), 900–901.

Glott, R., Schmidt, P., & Ghosh, R. (March 2010). Wikipedia survey—Overview of results: Collaborative Creativity Group, United Nations University—Maastricht Economic and Social Research and Training Centre on Innovation and Technology (UNU-MERIT). Retrieved from http://www.wikipediastudy.org/docs/Wikipedia_Overview_15March2010-FINAL.pdf.

Gould, C. C. (1988). *Information needs in the humanities: An assessment*. Stanford, CA: Research Libraries Group.

Gould, C. C., & Handler, M. J. (1989). *Information needs in the social sciences: An assessment*. Mountain View, CA: Research Libraries Group.

Gould, C. C., & Pearce, K. (1991). *Information needs in the sciences: An assessment*. Mountain View, CA: Research Libraries Group.

Grafstein, A. (2002). A discipline-based approach to information literacy. *Journal of Academic Librarianship, 28*(4), 197–204.

Gratch-Lindauer, B. (2002). Comparing the regional accreditation standards: Outcomes assessment and other trends. *Journal of Academic Librarianship, 28*(1/2), 14–25.

Gratch-Lindauer, B. (2007). Information literacy-related student behaviors: Results from the NSSE items. *College & Research Libraries News, 68*(7), 432–441.

Gratch-Lindauer, B., Arp, L., & Woodard, B. S. (2004). The three arenas of information literacy assessment. *Reference & User Services Quarterly, 44*(2), 122–129.

Green, R., & Bowser, M. (2006). Observations from the field. *Journal of Library Administration, 45*(1), 185–202.

Gross, M. (1995). The imposed query. *RQ: Research Quarterly, 35*(2), 236–243.

Gross, M. (1998). The imposed query: Implications for library service evaluation. *Reference & User Services Quarterly, 37*(3), 290–299.

Gross, M., & Latham, D. (2007). Attaining information literacy: An investigation of the relationship between skill level, self-estimates of skill, and library anxiety. *Library & Information Science Research, 29*(3), 332–353.

Hacker, B., & Stevens, R. (1975). Evaluating the AIMLO Project. Paper presented at the University of Denver Conference on the Evaluation of Library Instruction, Denver, CO.

Hacker, D., & Sommers, N. (2011). *A writer's reference*. Boston, Mass: Bedford/St. Martin's. Hamrick, F. A., Schuh, J. H., & Shelley, M. C. (2004). Predicting higher education graduation rates from institutional characteristics and resource allocation. *Education Policy Analysis Archives, 12*, 19–50.

Hardesty, L., Lovrich, N. P., & Mannon, J. (1982). Library-use instruction: Assessment of the long-term effects. *College & Research Libraries, 43*, 38–46.

Hardesty, L., Schmitt, J. P., & Tucker, J. M. (1986). *User instruction in academic libraries: A century of selected readings.* Metuchen, NJ: Scarecrow Press.

Hattie, J. (1999). Influences on student learning. Inaugural lecture: Professor of Education,

University of Auckland, given on August, 2, 1999. Retrieved January 7, 2013, from http://xn--www-rp0a.teacherstoolbox.co.uk/downloads/managers/Influencesonstudent.pdf

Hattie, J. (2005). What is the nature of evidence that makes a difference to learning? Paper presented at the *ACER Research Conference: 2005-Using data to support*

learning, Australian Council for Educational Research 7. Retrieved January 7, from http://research.acer.edu.au/cgi/viewcontent.cgi?article=1008&context=research_conference_2005.

Hauser, R. M., & Warren, J. R. (1996). Socioeconomic indexes for occupations: A review, update, and critique. CDE Working Paper No. 96-01. Madison: Center for Demography & Ecology, University of Wisconsin-Madison.

Hawthorne, J. (2008). Accountability and comparability: What's wrong with the VSA approach? *Liberal Education, 94*(2), 24.

Head, A. J., & Eisenberg, M. (2010). *Truth be told: How college students evaluate and use information in the digital age*. Seattle: Information School, University of Washington.

Hidi, S., & Boscolo, P. (2008). Motivation and writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 144–157). New York: Guilford Press.

Horton, F. W. (2008). *Understanding information literacy: A primer*. Paris: UNESCO. Retrieved January 7, 2013, from http://www.uis.unesco.org/Communication/Documents/157020E.pdf

Howard, R. M., Serviss, T., & Rodrigue, T. K. (2010). Writing from sources, writing from sentences. *Writing & Pedagogy, 2*(2), 177–192.

Huot, B. (1996). Toward a new theory of writing assessment. *College Composition & Communication, 47*(4), 549–566.

Institute of Museum and Library Services (IMLS). (2012). RAILS: Rubric Assessment of Information Literacy Skills. Retrieved December 15, 2012, from www.railsontrack.info.

International ICT Literacy Panel. (2002). *Digital transformation: A framework for ICT literacy*. Princeton, NJ: Educational Testing Service.

Ivanitskaya, L., DuFord, S., Craig, M., & Casey, A. M. (2008). How does a pre-assessment of off-campus students' information literacy affect the effectiveness of library instruction? *Journal of Library Administration, 48*(3–4), 509–525.

Ivanitskaya, L., Laus, R., & Casey, A. M. (2005). Research Readiness Self-Assessment. *Journal of Library Administration, 41*(1), 167–183.

Jackson, P. A. (2006). Plagiarism instruction online: assessing undergraduate students' ability to avoid plagiarism. *College & Research Libraries, 67*(5), 418–428.

Jackson, R. (2007). Cognitive development: The missing link in teaching information literacy skills. *Reference & User Services Quarterly, 46*(4), 28–32.

Jacobson, T., & Xu, L. (2002). Motivating students in credit-based information literacy courses: Theories and practice. *Portal: Libraries & the Academy, 2*(3), 423–441.

Johnson, C. S. (2006). The Analytic assessment of online portfolios in technical communication: a model. *Journal of Engineering Education*, 95(4), 279-287.

Johnson, C. S., & Elliot, N. (Fall 2004). Portfolio assessment in program evaluation. Paper presented at the Proceedings of the American Society for Engineering Education, Middle Atlantic Section Conference. Retrieved from http://web.njit.edu/~elliot/presentations/asee04.pdf.

Jullien, N. (2012, May 7). "What We Know About Wikipedia. A Review of the Literature Analyzing the Project(s)." *Social Science Research Network*. Papers.ssrn.com.papers.ssrn.com.

Kanter, S. B. (2006). *Embodying research: A study of student engagement in research writing.* Ph.D. diss., Indiana University of Pennsylvania.

Katz, I. (2007). Testing information literacy in digital environments: ETS's iSkills assessment. *Information Technology & Libraries, 26*(3), 3–12.

Katz, I., Elliot, N., Attali, Y., Scharf, D., Powers, D., Huey, H., . . . Briller, V. (2008). The assessment of information literacy: A case study (Vol. ETS RR-08-33). Princeton, NJ: Educational Testing Service.

Keene, J., Colvin, J., & Sissons, J. (2010). Mapping student information literacy activity against Bloom's taxonomy of cognitive skills. *Journal of Information Literacy, 4*(1), 6–21.

Keller, J. M. (1987). The systematic process of motivational design. *Performance + Instruction, 26*(9–10), 1–8. doi: 10.1002/pfi.4160260902.

Kelly, M. C. (1995). Student retention and academic libraries. *College & Research Libraries News, 56*(11), 757–759.

Kirk, T. (1971). A comparison of two methods of library instruction for students in introductory biology. *College & Research Libraries, 32*(6), 465–474.

Kirk, T. (1975). Bibliographic instruction: A review of research. Paper presented at the University of Denver, Conference on the Evaluation of Library Instruction, Denver, CO.

Kittur, A., Chi, E., Pendelton, B., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. Proceedings of CHI '07, CHI Conference on Human Factors in Computing Systems. San Jose, CA, USA — April 30 - May 03, 2007. New York: ACM.

Knapp, P. B. (1966). *The Monteith College library experiment*. New York: Scarecrow Press.

Knight, L. A. (2006). Using rubrics to assess information literacy *Reference Services Review, 34*(1), 43–55.

Knowles, M. S., Holton, E. F., & Swanson, R. A. (2005). *The adult learner: The definitive classic in adult education and human resource development*. Amsterdam, The Netherlands: Elsevier.

Kobrin, J. L., Deng, H., & Shaw, E. J. (2007). Does quantity equal quality? The relationship between length of response and scores on the SAT essay. *Journal of Applied Testing Technology, 8*(1), 1–15.

Kohl, D. F., & Wilson, L. (1986). Effectiveness of course-integrated bibliographic instruction in improving coursework. *RQ: Research Quarterly, 27*(2), 206–211.

Konieczny, P. (2007) "Wikis and Wikipedia as a Teaching Tool. *International Journal of Instructional Technology and Distance Learning* . 4(1).

Konieczny, P. (2012). Wikis and Wikipedia as a Teaching Tool: Five Years Later. *First Monday* 17(9-3)

Koufogiannakis, D., & Wiebe, N. (2006). Effective methods for teaching information literacy skills to undergraduate students: A systematic review and meta-analysis. *Evidence Based Library & Information Practice, 1*(3): 3-43.

Kramer, L. A., & Kramer, M. B. (1968). The college library and the drop-out. *College & Research Libraries*, 29(4): 310–312.

Krathwohl, D. R. (1998). *Methods of educational and social science research: An integrated approach* (2nd ed.). Long Grive, IL: Waveland Press.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212–218.

Kuh, G., & Gonyea, R. (2003). The role of the academic library in promoting student engagement in learning. *College & Research Libraries, 64*(4), 256–282.

Kuh, G., Kinzie, J., Schuh, J., & Whitt, E. (2010). *Student success in college: Creating conditions that matter*. San Francisco, CA: Jossey-Bass.

Kuhlthau, C. C. (1988). Developing a model of the library search process: Cognitive and affective aspects. *RQ: Research Quarterly, 28*, 232–242.

Kuhlthau, C. C., Caspari, A. K., & Maniotes, L. K. (2007). *Guided inquiry: Learning in the 21st century*. Westport, CT: Libraries Unlimited.

Kurbanoglu, S. S., Akkoyunlu, B., & Umay, A. (2006). Developing the information literacy self-efficacy scale. *Journal of Documentation, 62*(6), 730–743.

Kuteeva, M. (2011). Wikis and academic writing: Changing the writer–reader relationship. *English for Specific Purposes, 30*(1), 44–57.

Kwon, N. (2008). A mixed-methods investigation of the relationship between critical thinking and library anxiety among undergraduate students in their information search process. *College & Research Libraries, 69*(2), 117–131.

Kwon, N., Onwuegbuzie, A. J., & Alexander, L. (2007). Critical thinking disposition and library anxiety: Affective domains on the space of information seeking and use in academic libraries. *College & Research Libraries, 68*(3), 268–278.

Larson, R. L. (1982). The "research paper" in the writing course: A non-form of writing. *College English, 44*(8), 811–816.

Lazarow, M. (2007). The evidence-based model of information literacy research: A critique. In S. Lipu, K. Williamson, & A. Lloyd (Eds.), *Exploring methods in information literacy research* (pp. 171–184). Wagga Wagga, Australia: Centre for Information Studies, Charles Sturt University.

Leckie, G. J. (1996). Desperately seeking citations: Uncovering faculty assumptions about the undergraduate research process. *Journal of Academic Librarianship, 22*(3), 201–208.

LibQual+. (2009). About the LibQual survey. Retrieved February 22, 2011, from http://www.libqual.org/Information/index.cfm.

Lih, A. (2009). *The Wikipedia revolution: How a bunch of nobodies created the world's greatest encyclopedia*. New York: Hyperion.

Limberg, L., & Sundin, O. (2006). Teaching information seeking: Relating information literacy education to theories of information behaviour. *Information Research, 12*(1), 9.

Lloyd, A., & Williamson, K. (2008). Towards an understanding of information literacy in context: Implications for research. *Journal of Librarianship & Information Science, 40*(1), 3–12.

Macrorie, K. (1988). *The I-Search paper*. Portsmouth, NH: Heinemann

Macrorie, K. (1988). *The I-Search paper: Revised Edition of "Searching Writing.*" Portsmouth, NH: Heinemann Educational Books.

Mallinckrodt, B., & Sedlacek, W. E. (1987). Student retention and the use of campus facilities by race. *NASPA Journal, 24*(3), 28–32. doi: EJ355278 (ERIC).

Mark, A. E., & Boruff-Jones, P. D. (2003). Information literacy and student engagement: What the National Survey of Student Engagement reveals about your campus. *College & Research Libraries, 64*(6), 480–493.

Martin, J. (2011). Investigation of factors affecting information literacy student learning outcomes fails to undercover significant findings. *Evidence Based Library & Information Practice, 6*(2), 59–60.

Marton, F., & Säljö, R. (1976). On qualitative diferences in learning: I—Outcome and Process. *British Journal of Educational Psychology, 46*(1), 4–11.

Mathews, W. (1877). *Hours with men and books*. Chicago: Griggs.

McClure, R., Cooke, R., & Carlin, A. (2011). The search for the skunk ape: Studying the impact of an online information literacy tutorial on student writing. *Journal of Information Literacy, 5*(2), 26–45.

McCurry, D. (2010). Can machine scoring deal with broad and open writing tests as well as human readers? *Assessing Writing, 15*(2), 118–129. doi: 10.1016/j.asw.2010.04.002.

McGuinness, C. B., M. . (2007). Using reflective journals to assess the research process. *Reference Services Review, 35*(1), 21–40.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online*

learning studies. (Contract number ED-04-CO-0040 Task 0006). Office of Planning, Evaluation, and Policy Development Policy-Program Studies Service Center for Technology in Learning. Washington, DC: U.S. Department of Education.

Mellon, C. A. (1986). Library anxiety: A grounded theory and its development. *College & Research Libraries, 47*, 160–165.

Mellon, C. A., & Sass, E. (1981). Perry and Piaget: Theoretical framework for effective college course development. *Educational Technology, 21*(5), 29–33.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13.

Mezick, E. (2007). Return on investment: Libraries and student retention. *Journal of Academic Librarianship, 33*(5), 561–566.

Middaugh, M. F. (2010). *Planning and assessment in higher education: Demonstrating institutional effectiveness* (Vol. 1). San Francisco: Jossey-Bass.

Middle-States. (2003). *Developing research & communication skills: Guidelines for information literacy in the curriculum*. Philadelphia, PA: Middle States Commission on Higher Education.

Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation, 7*(3). Retrieved January 7, 2013, from http://PAREonline.net/getvn.asp?v=7&n=3

Mulherrin, E., & Abdul-Hamid, H. (2010). The evolution of a testing tool for measuring undergraduate information literacy skills in the online environment. *Communications in Information Literacy, 3*(2), 204.

Nahl, D. (1996). *Affective monitoring of Internet learners: Perceived self-efficacy and success*. Proceedings of the Annual Meeting of the American Society for Information Science (ASIS). (v. 33, pp. 100-109).

Nahl, D., & Bilal, D. (2007). *Information and emotion: The emergent affective paradigm in information behavior research and theory*. Medford, NJ: Information Today.

National Center for Education Statistics. IPEDS (Integrated Post Secondary Education Data System): Library Statistics Program. Retrieved April 26, 2012, from http://nces.ed.gov/surveys/libraries/.

National Council of Teachers of English NCTE. (1983). Resolution on writing across the curriculum. *NCTE Position Statement*. 1983 Annual Business Meeting, Denver, CO.

Neely, T. Y. (2000). *Aspects of information literacy: A sociological and psychological study.* Ph.D. diss., University of Pittsburgh, PA. Retrieved from http://proquest.umi.com/pqdweb?did=732062851&Fmt=7&clientId=16246&RQT=309&VName=PQD.

Neely, T. Y. (2006). *Information literacy assessment: Standards-based tools and assignments*. Chicago: American Library Association.

New Jersey Institute of Technology. (May 2009). Institute Information Literacy Plan. Newark, NJ. Retrieved January 7, 2013 from http://library.njit.edu/library/docs/njit-info-lit-plan-caa-approved-05-20-2009.pdf .

New Jersey Institute of Technology. Institutional Research and Planning. (2012). *NJIT Factbook* Retrieved May 29, 2012, Retrieved from http://www.njit.edu/irp/factbook/.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175.

Niedbala, M. A., & Fogleman, J. (2010). Taking library 2.0 to the next level: Using a course wiki for teaching information literacy to honors students. (Special issue). *The 14th Off-Campus Library Services Conference Proceedings, part 2, 50*(7/8), 867–882. doi: 10.1080/01930826.2010.488986.

Nielsen, F. A. (2007). Scientific citations in Wikipedia. *First Monday, 12*(8). Retrieved January 7, 2013, from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1997/1872

Norgaard, R. (2003). Writing information literacy: Contributions to a concept. *Reference & User Services Quarterly, 43*, 124–130.

O'Connor, L. G. (2006). *Librarians' professional struggles in the information age: A critical analysis of information literacy.* Ph.D. diss., Kent State University. Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=kent1153761756.

O'Connor, L. G., Radcliff, C. J., & Gedeon, J. A. (2001). Assessing information literacy skills: Developing a standardized instrument for institutional and longitudinal measurement. Paper presented at the Crossing the Divide: Proceedings of the 10th National Conference of the Association of College and Research Libraries, Chicago, IL.

O'Neill, P., Moore, C., & Huot, B. A. (2009). *A Guide to College Writing Assessment*: Utah State University Press.

O'Sullivan, D. (2009). *Wikipedia: A new community of practice?* Burlington, VT: Ashgate.

Oakleaf, M. (2006). *Assessing information literacy skills: A rubric approach.* Ph.D. diss., University of North Carolina at Chapel Hill.

Oakleaf, M. (2007). Using rubrics to collect evidence for decision-making: What do librarians need to learn? *Evidence Based Library & Information Practice, 2*(3), 27.

Oakleaf, M. (2008). Dangers and opportunities: A conceptual map of information literacy assessment approaches. *Portal: Libraries & the Academy, 8*(3), 233–253.

Oakleaf, M. (2010). The value of academic libraries: A comprehensive research review and report. Chicago: Association of College & Research Libraries.

Ondrusek, A. (2008). Information literacy. In M. L. Radford & P. Snelson (Eds.), *Academic library research: Perspectives and Current Trends* (pp. 48–81). Chicago: Association of College & Research Libraries.

Ondrusek, A., Dent, V., & Bonadie, J. (2005). A longitudinal study of the development and evaluation of an information literacy test. *Reference Services Review, 33*(4), 388–417.

Pajares, F., & Valiante, G. (2008). Self-efficacy beliefs and motivation in writing development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 144–157). New York: Guilford Press.

Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Francisco: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students : A third decade of research* (Vol. 1). San Francisco: Jossey-Bass.

Pedhazur, E., & Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Perelman, L. (2005, May 29). New SAT: Write long, badly, and prosper. *Los Angeles Times.* Retrieved January 7, 2013, http://articles.latimes.com/2005/may/29/opinion/oe-perelman29.

Perkins, F. B. (1876). On professorships of books and reading. In U.S. Bureau of Education (Ed.), *Public libraries in the United States of America* (pp. 230–239). Washington, DC: U.S. Government Printing Office.

Pinto, M. (2009). Design of the IL-HUMASS survey on information literacy in higher education: A self-assessment approach. *Journal of Information Science, 20*(10), 1–18.

Project SAILS (Standardized Assessment of Information Literacy Skills). (2011). Project SAILS skill sets for the 2011–2012 academic year: Using ACRL information literacy standards Retrieved April 20, 2012, from https://www.projectsails.org/SkillSets.

Quaratiello, A. R., & Devine, J. (2011). *The college student's research companion: Finding, evaluating, and citing the resources you need to succeed* (5th ed.). New York: Neal-Schuman.

Radcliff, C. J., Jensen, M. L., Salem, J. A., Burhanna, K. J., & Gedeon, J. A. (2007). *A practical guide to information literacy assessment for academic librarians*. Westport, CT: Libraries Unlimited.

Reason, R. D. (2009). An examination of persistence research through the lens of a comprehensive conceptual framework. *Journal of College Student Development, 50*(6), 659–682.

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*(4), 435–448. doi: 10.1080/02602930902862859.

Ritter, K. (2005). The economics of authorship: Online paper mills, student writers, and first-year composition. *College Composition & Communication, 56*(4), 601–631.

Roberts, J. D. (2004). Senior student nurses information seeking skills: a comparative study. *Nurse Education Today, 24*(3), 211–218.

Rosenblatt, S. (2010). They can find it but they don't know what to do with it: Describing the use of scholarly literature by undergraduate students. *Journal of Information Literacy, 4*(2), 50–61.

Rosenshine, B., & Meister, C. (1992). The use of scaffolds for teaching higher-level cognitive strategies. *Educational Leadership, 49*(7), 26–33.

Rushing, D., & Poole, D. (2002). The role of the library in student retention. In Kelly, M. & A. Kross (Eds.), *Making the grade: Academic libraries and student success* (pp. 91–101). Chicago: American Library Association.

Ruszkiewicz, J. J. (2011). *The Scott, Foresman handbook for writers*. Boston, Mass: Longman. Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin, 135*(1), 1–22. doi: 10.1037/a0013978.

Samson, S. (2010). Information literacy learning outcomes and student success. *Journal of Academic Librarianship, 36*(3), 202–210.

Saunders, E. S. (2010). *Information literacy as a student learning outcome: As viewed from the perspective of institutional accreditation.* Ph.D. diss., Simmons College. Retrieved from http://dspace.nitle.org/handle/10090/20894.

Saunders, L. (2007). Regional accreditation organizations' treatment of information literacy: Definitions, collaboration, and assessment. *Journal of Academic Librarianship, 33*(3), 317–326.

Saunders, L. (2009). The future of information literacy in academic libraries: A Delphi study. *Portal: Libraries & the Academy, 9*(1), 99–114.

Scharf, D., Elliot, N., Huey, H., Briller, V., & Joshi, K. (2007). Direct assessment of information literacy using writing portfolios. *Journal of Academic Librarianship, 34*, 462–477.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs: A think tank white paper*. Washington, DC: American Educational Research Association.

Schunk, D. H. (1995). Self-efficacy and education and instruction. In J. E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment: Theory, research, and application* (pp. 281–303). New York: Plenum Press.

Schunk, D. H., & Zimmerman, B. J. (1994). *Self-regulation of learning and performance: Issues and educational applications*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Selegean, J., Thomas, M., & Richman, M. (1983). Long-range effectiveness of library use instruction. *College & Research Libraries, 44*(6), 476–480.

Sen, B. A. (2010). Reflective writing: a management skill. *Library management, 31*(1/2), 79–93.

Shadish, W., Cook, T., & Leviton, L. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.

Shapiro, J., & Hughes, S. (1996). Information literacy as a liberal art? *Educom Review, 31*, 31–35.

Shavelson, R. (2009). *Measuring college learning responsibly: Accountability in a new era*. Palo Alto, CA: Stanford University Press.

Sherman, M., Martin, J. A., & An, X. (2011). The impact of library instruction on the quality of student project performance in an advanced financial management case class. *Journal of Business & Finance Librarianship, 17*(1), 51–76. doi: 10.1080/08963568.2012.630646.

Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417.

Smagorinsky, P. (2006). *Research on composition: Multiple perspectives on two decades of change*. New York: Teachers College Press.

Snavely, L. L., & Wright, C. A. (2003). Research portfolio use in undergraduate honors education: Assessment tool and model for future work. *Journal of Academic Librarianship, 29*(5), 298–303.

Society of College National & University Libraries (SCONUL). (2004). Learning outcomes and information literacy. London: Higher Education Academy.

Sonley, V., Turner, D., Myer, S., & Cotton, Y. (2007). Information literacy assessment by portfolio: A case study. *Reference Services Review, 35*(1), 41–70.

Suskie, L. (2009). *Assessing student learning: A common sense guide*. San Franciso: Jossey-Bass.

Tanilon, J., Vedder, P., Segers, M., & Tillema, H. (2011). Incremental validity of a performance-based test over and above conventional academic predictors. *Learning & Individual Differences, 21*(2), 223–226. doi: DOI: 10.1016/j.lindif.2010.12.005.

Tanni, M., & Sormunen, E. (2008). A critical review of research on information behavior in assigned learning tasks. *Journal of Documentation, 64*(6), 893–914. doi: 10.1108/00220410810912442.

Thornton-Verma, H. (2012, April 18). "Reaching the Wikipedia Generation: Reference Roundtable Tackles Trends and Thorny Issues." *Library Journal.* lj.libraryjournal.com.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89–125.

Trochim, W. M. K. (2001). *The research methods knowledge base* (2nd ed.). Cincinnati, OH: Atomic Dog.

Tucker, J. M. (1980). Articles on library instruction in colleges and universities, 1876–1932 *Occasional Papers*. Champaign, IL: Graduate School of Library and Information Science. University of Illinois at Urbana-Champaign. (Reprinted from: ERIC ED 187 330).

Tyler, R. W. (1980). Landmarks in the Literature: What Was Learned from the Eight-Year Study. New York University Education Quarterly, 11(2), 29-32.

Tyler, R. W. (1942). *Eight year study*. Progressive Education Association.

Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.

UCLA Library Information Literacy Program Steering Committee. (2005). *Information literacy at UCLA: The core competencies*. Los Angeles: University of California, Los Angeles Library. Retrieved from http://escholarship.org/uc/item/8kh5v4q0.

van Helvoort, J. (2010). A scoring rubric for performance assessment of information literacy in Dutch higher education. *Journal of Information Literacy, 4*(1), 22–39.

Vygotsky. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*(8), 19–27.

Walsh, A. (2009). Information literacy assessment: Where do we start? *Journal of Librarianship & Information Science, 41*(1), 19–28.

Walton, M., & Archer, A. (2004). The Web and information literacy: Scaffolding the use of Web sources in a project-based curriculum. *British Journal of Educational Technology, 35*(2), 173–186.

Warner, D. (2008). *A disciplinary blueprint for the assessment of information literacy*. Westport, CT: Libraries Unlimited.

Warriner, J. E. (1957). *English grammar and composition: A complete handbook*. New York: Harcourt, Brace.

Weiner, J. M. (2011). Is there a difference between critical thinking and information literacy? A systematic review 2000–2009. *Journal of Information Literacy, 5*(2), 81–92.

Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Werking, R. H. (1980). Evaluating bibliographic education: A review and critique. *Library Trends, 29*(1), 160–161.

White, E. M. (1995). An apologia for the timed impromptu essay test. *College Composition & Communication, 46*(1), 30–45.

White, E. M. (2005). The scoring of writing portfolios: Phase 2. *College Composition & Communication, 56*(4), 581–600.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461–481.

Whitmire, E. (2001). *The influence of academic library experience on undergraduates' critical thinking during three years of college.* Ph.D. diss., University of Michigan, Ann Arbor.

Whitmire, E. (2002a). Academic library performance measures and undergraduates' library use and educational outcomes. *Library & Information Science Research, 24*(2), 107–128.

Whitmire, E. (2002b). Disciplinary differences and undergraduates' information-seeking behavior. *Journal of the American Society for Information Science & Technology, 53*(8), 631–638.

Whitmire, E. (2003). Epistemological beliefs and the information-seeking behavior of undergraduates. *Library & Information Science Research, 25*(2), 127–142.

Whitmire, E. (2006). African-American undergraduates and the university academic library. *Journal of Negro Education, 75*(1), 60–66.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 70*(9), 703–713.

Wiggins, G. (1990). *The case for authentic assessment*. ERIC Digest. ED328611. Retrieved from http://www.eric.ed.gov/PDFS/ED328611.pdf.

Wiggins, G. (1994). The constant danger of sacrificing validity to reliability: Making writing assessment serve writers. *Assessing Writing, 1*(1), 129–139.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.

Wikimedia Foundation. (June 2012). Case studies: How professors are teaching with Wikipedia. Wikipedia Education Program. Retrieved September 15, 2012. http://outreach.wikimedia.org/wiki/Education/Case_Studies

Wikimedia Foundation. (2012). Wikipedia as a teaching tool (Bookshelf). *Wikipedia.* Retrieved June 25, 2012, from http://outreach.wikimedia.org/w/index.php?title=Wikipedia_as_a_Teaching_Tool_(Bookshelf)&oldid=8392.

. Wikipedia Education Program. (2012). *Wikipedia.* Retrieved June 25, 2012, from http://outreach.wikimedia.org/w/index.php?title=Wikipedia_Education_Program&oldid=37065.

Wikipedia. Frequently asked questions. (2012, June 3). Retrieved June 15, 2012, from http://wikimediafoundation.org/w/index.php?title=FAQ/en&oldid=82148

. Wikipedia. (2012a). *Wikipedia* Retrieved June 15, 2012, from http://en.wikipedia.org/w/index.php?title=Wikipedia&oldid=497610519

. Wikipedia. (2012b). Policies and guidelines. Retrieved June 15, 2012, from http://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines.

. Wikipedia. (2012c). School and university projects. Retrieved June 25, 2012, from http://en.wikipedia.org/wiki/Wikipedia:School_and_university_projects.

Wilson, C. (2008). The wisdom of the chaperones: Digg, Wikipedia, and the myth of Web 2.0 democracy. *Slate.com*. Retrieved from http://www.slate.com/articles/technology/technology/2008/02/the_wisdom_of_the_chaperones.html.

Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation, 37*(1), 3–15.

Winsor, J. (1880). *The college library*. Washington, DC: U.S. Government Printing Office.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology & Psychiatry, 17*(2), 89–100.

Wriston, H. M. (1959). *Academic procession: Reflections of a college president*. New York: Columbia University Press.

Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition & Communication, 50*(3), 483–503.

Young, J. (2007). *Validity of the measure of academic proficiency and progress (MAPP).* Princeton, NJ: Educational Testing Service.

Zehner, D. C. B. (2010). *Factors affecting information literacy perception and performance.* Ed.D. diss., University of South Carolina, Columbia.

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*(4), 660–668.

Zimmerman, B. J., Schunk, D. H., & Martin, J. (2004). Educational psychology: A century of contributions. *Contemporary Psychology, 49*(6), 722.

Zurkowski, P. (1974). *The information service environment: Relationships and priorities*. . (ERIC Clearinghouse on Information Resources. ED 100391).Washington DC: National Commission on Libraries and Information Science.

`