Development And Validation Of The Beile Test Of Information Literacy For Education (b-tiled)

2005

Penny Beile O'Neil
University of Central Florida

DEVELOPMENT AND VALIDATION OF THE BEILE TEST OF INFORMATION LITERACY FOR EDUCATION (B-TILED)

by

PENNY M. BEILE O'NEIL
B.A. University of Kentucky, 1987
M.A. University of Kentucky, 1989
M.S. University of Kentucky, 1991

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Educational Studies
in the College of Education
at the University of Central Florida
Orlando, Florida

Fall Term
2005

Major Professor: David N. Boote

ABSTRACT

Few constituencies exist where it is more important to produce information literate individuals than teacher candidates, yet rarely is it suggested that practitioners entering the field are adequately prepared to teach and model information literacy to their students. As a result, information literacy has been established as a key outcome by a number of teacher education accrediting bodies and professional associations. Corollary to this initiative is the effort to develop valid instruments that assess information literacy skills. Yet, at the time of this dissertation, no rigorously reviewed instruments were uncovered that measure the information literacy skills levels of teacher candidates.

The study describes the development and validation of the Beile Test of Information Literacy for Education (B-TILED). Funded in part by the Institute for Library and Information Literacy Education and the Institute of Museum and Library Services, the study is part of a national initiative spear-headed by the Project for the Standardized Assessment of Information Literacy Skills (SAILS).

Test content is based on nationally recognized standards from the International Society for Technology in Education and the Association of College and Research Libraries. Procedures designed to enhance the scale's validity were woven throughout its development. 172 teacher education students at a large, metropolitan university completed a protocol consisting of 22 test items and 13 demographic and self-percept items. This instrument can be used to inform curricular and instructional decisions and to provide evidence of institutional effectiveness for program reviews.

Henry Alexander Minton
When faced with the choice, you thought a book in hand more important than food in the belly. You are, and always will be, my hero.

Overton (Toby) Kavanaugh
You affected me without ever quite realizing how deeply. I hope you have found your kingdom by the sea. A day doesn't pass that I don't think of you.

Kelly Anne Beile
In your own inimitable way you encouraged me to follow my bliss. My wish for you is peace, love, and happiness in this lifetime. You are the reason I was born.

Thomas O'Neil
You opened to me a world with endless possibilities. I would never have accomplished this goal without your support. I love you Porgy, I got my man.

The people of New Orleans, Southeast Louisiana, and the Gulf Coast
You are forever in my heart. Laissez les bon temps roulez.

Chorus I

Where is the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?

- - T.S. Eliot, 1934
Choruses from "The Rock"

Upon this Age

Upon this gifted age, in its dark hour,
Rains from the sky a meteoric shower
Of facts… they lie unquestioned, uncombined.
Wisdom enough to leech us of our ill
Is daily spun; but there exists no loom
To weave it into fabric...

- - Edna St. Vincent Millay, 1939
Huntsman, What Quarry?

The fragmentation of rational knowledge in the postmodern world has produced a focus on information that is unaware of its history.

- - Marcus Green

ACKNOWLEDGMENTS

Numerous individuals provided support and assistance throughout the course of this study. I am grateful to all of those who contributed to this project.

Special thanks go to the team from the Project for the Standardized Assessment of Information Literacy Skills (Project SAILS), who provided support and training for initial item development, one-on-one testing, and small group testing. My sincere thanks are extended to Dr. Lisa O'Connor, Carolyn Radcliff, Dr. Julie Gedeon, Mary Thompson, and Rick Wiggins. My appreciation is also extended to Dr. Gordon Taub, who comprised the other half of the UCF project team to develop education-specific information literacy test items.

This project was supported, in part, by a Project SAILS-awarded fellowship funded by the Institute for Library and Information Literacy Education (ILILE). I wish to acknowledge the contributions Project SAILS, ILILE, and the Institute of Museum and Library Services (IMLS) have made towards furthering information literacy assessment.

Assistance from numerous other individuals also contributed to the success of this project. I am indebted to five leading experts in the field of information literacy instruction for education students. Many, many thanks to JoAnn Carr, University of Wisconsin – Madison; Natasha (Tasha) Cooper, Syracuse University; Francine DeFranco, University of Connecticut; Dr. Claudia Morner, University of New Hampshire; and Dr. Scott Walter, University of Kansas, who reviewed test items for content, clarity, accuracy, and institutional objectivity. This study relied heavily upon Claudia Morner's 1993 dissertation, and thanks are extended to Dr. Morner for allowing

me to adapt some of her test items. Myrtle Harris, University of the West Indies – Mona, Jamaica campus, also deserves recognition for her critical reading and comments in regard to the items and assistance with student one-on-one testing. Each person played a significant role in the construction of this test.

As always, thanks to the UCF Libraries Interlibrary Loan department for their efficient and capable service in procuring information not locally held.

My experience throughout this project was enhanced by my fellow doctoral students. I would like to thank the following doctors and soon-to-be doctors for making this an enjoyable experience: Thomas Fisher, Robert Duggan, Cathy Freytag, Stacy DeZutter, Nehemiah (Nammie) Ichilov, Nancy Cummings, Stacey Keller, Becky Fiedler, Julie Phelps, Amy Scheik, and Debbie Wheeldon. Of course, this list would not be complete without mention of a truly inspirational doctoral candidate, library colleague, and my best bud, Donna Goda.

Recognition and thanks also to the following programs and policies, and to the individuals who implement and represent them: the University of Central Florida, without their tuition waiver program it is highly unlikely my dream of a doctorate would ever be made manifest; the UCF Office of Graduate Studies, for their personal interest and research presentation opportunities; the UCF Libraries, for approving my professional development leave request and support in general; and the UCF College of Education, for finding my plea to "be a better librarian" acceptable fodder for a Ph.D. hopeful. Effervescent and always patient, Virenda Barbour smoothed numerous bumps in the road. Several faculty members not on my dissertation committee also (statistically?) significantly impacted my academic and professional growth. Dr. Lea

Witta, Dr. Stephen Sivo, and Dr. Laura Blasi, your time, teaching, and insights challenged me to think, produce, and contribute at a higher level.

Finally, I would like to thank my committee, the dream team - - Dr. Linda Alexander, University of South Florida, unconventional, independent and supportive, your determination to achieve your academic goals has been an inspiration to me; Dr. Donna Baumbach, SUNLINK, a visionary professional and hard-workin' woman, your efforts to connect and support Florida media specialists, school libraries, and students have, and will continue to have, a profound and lasting impact on the state; Dr. Charles (Chuck) Dziuban, University of Central Florida, a scholar in the truest sense of the word, you have successfully balanced a distinguished academic career with family life and civic responsibilities, while still managing to get in a little fly-fishing time; Dr. Jeff Kaplan, University of Central Florida, reflective, insightful, and above all empathetic, you are the calm in the storm, the person we most depend upon to get the job done.

Committee chair and faculty mentor, Dr. David Boote, University of Central Florida, talented academician, a professor dedicated to his students, and the embodiment of the spirit of intellectual inquiry. I attribute my academic success and level of professional activity to your individual support and attention and your expectation of nothing less than my best. For this, I thank you. I feel extremely fortunate to have had the opportunity to work with such a unique and talented group of individuals and have learned many valuable lessons from each of you. You are remarkable faculty. I wish all doctoral students the same experience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

AACU	Association of American Colleges and Universities
AAHE	American Association for Higher Education
AASL	American Association of School Librarians
ACRL	Association of College and Research Libraries
ACS	Associated Colleges of the South
AECT	Association of Educational and Communications Technology
ALA	American Library Association
APA	American Psychological Association
B-TILED	Beile Test of Information Literacy for Education
EBSS	Education and Behavioral Sciences Section
EPIC	Electronic Publishing Initiative at Columbia
ETS	Educational Testing Service
ILILE	Institute of Library and Information Literacy Education
IMLS	Institute of Museum and Library Services
INTASC	Interstate New Teacher Assessment and Support Consortium
ISTE	International Society for Technology in Education
MSCHE	Middle States Commission on Higher Education
NCATE	National Council for the Accreditation of Teacher Education
NEASC	New England Association of Schools and Colleges
NETS*T	National Educational Technology Standards for Teachers
NWCCU	Northwest Commission on Colleges and Universities

OEAS	Operational Excellence and Assessment Support, at UCF
Project SAILS	Project for the Standardized Assessment of Information Literacy Skills
UCF	University of Central Florida

CHAPTER ONE
INTRODUCTION

This paper chronicles the development and validation of an information literacy skills test for education students. Information literacy is an evolving topic and only lately have efforts been expended to develop standardized tests for assessing students' information literacy skills levels as they relate to a general liberal arts education. Even more recent is the move toward developing discipline-specific information literacy assessment measures. The subject of this study, the Beile Test of Information Literacy for Education (B-TILED), is designed to assess the information literacy skill levels of education students. As such, it is unique among information literacy assessment instruments.

Under the leadership of the Project for the Standardized Assessment of Information Literacy Skills (Project SAILS), an Institute of Museum and Library Services (IMLS)-funded enterprise housed at Kent State University, four teams of faculty from across the country were competitively chosen for a fellowship to populate a test item bank for the disciplines of biology, communications, education, and history. A project team from the University of Central Florida (UCF), comprised of Gordon Taub and the author, were selected to develop test items for education. The Institute for Library and Information Literacy Education (ILILE) provided additional funding for the education fellowship. The first part of the study explains Project SAILS-supported efforts, while the second part describes subsequent steps to develop a single form information literacy skills test for education.

The remainder of this chapter focuses on the destabilization of the traditional scholarly information model, and how the resulting flux has thrust the topic of information literacy to the forefront of academia. It is this phenomenon, along with the burgeoning interest in information literacy in the academy and the author's own professional interests, that have motivated and formed this study. Goals and scope of the study are described, and definitions of study terms and literature review descriptors follow. The chapter closes with a brief overview of the remaining chapters.

Background

In *The Name of the Rose*, Umberto Eco (1983) writes, "The Library defends itself, immeasurable as the truth it houses, deceitful as the falsehood it preserves. A spiritual labyrinth, it is also a terrestrial labyrinth" (p. 38). All who have wandered mesmerized through a seemingly endless maze of books understand Eco's reference to the earthly. However, the compelling spiritual nature of the library to which Eco alludes is less apparent. Is this appeal a function of the building as temple? Does Eco grant transcendent status to the library as repository of civilization's knowledge? Perhaps, instead, it is simply his tribute to its attempt at bridling the unruliness of the written record.

Because Eco (1983) was composing a murder mystery, he most likely refers to the winding library stacks as a probable location for mischief when he adds, "You might enter, and you might not emerge" (p. 38). One wonders if Eco was also presciently referring to the Internet, for it is precisely this development that has transformed his

portrayal of the quest for knowledge from the cogent to the chaotic, from the organized to the anarchic. The intent of this analogy is not to imply that information retrieved from the Internet is valueless, but that it has destabilized conventional information-seeking behavior in the academy. As the following example illustrates, the inarguable fact is that the Internet has forever altered how information is communicated.

John Derek de Solla Price, who studied the volume and growth rate of scientific information over time, hypothesized that the scientific literature grows at an exponential rate. Based on his prediction that scientific publication doubles in size every 10 to 15 years, de Solla Price concluded that it was no longer possible for an individual to attain mastery of the literature in any one branch of science (Erlendsson, 2003). With the emergence of the Internet as both a publishing venue and facilitator to information access and dissemination, de Solla Price's publication production estimations have been replaced by a new information explosion model. A central figure in this field is Hal Varian (2003), dean of the School of Information Management and Systems at the University of California at Berkeley, who recently reported on how much information is being generated annually.

Based on a 2002 survey, Varian (2003) found that the volume of information produced annually is measured in terms of exabytes,[1] and that 92% of information produced is in electronic format. To illustrate the magnitude of this amount of information, Varian offers the following example: if the 17 million books in the Library

---

[1] An exabyte is roughly equal to 1,000,000,000,000,000,000 (10^{18}) bytes.

of Congress were digitized, it would take 136 terabytes[2] to store the data. Conversely, if one year's worth of produced information was scanned, it would take 37,000 libraries the size of the Library of Congress to store it in print format. Varian goes on to add that new information is doubling at the rate of every two to three years.

Comparing de Solla Price's prediction regarding the proliferation of scientific publication with Varian's study of digital information is analogous to comparing apples with oranges; as of 2001, only about 8% of scholarly journals were on the Internet (Herring, 2001). Although both researchers investigated the quantity of information produced, de Solla Price focused on scientific publication while Varian researched general information production. Varian affirms that the Internet is the fastest growing communication medium of all time, and that the bulk of Internet generated information is electronic mail, Web page production, and instant messaging. The point of this discussion is that information produced today is not generally considered as reputable as the information that de Solla Price studied, leading Varian (2003) to surmise "information management is the challenge of the next 20 years" (p. 1).

Validating Varian's conclusions are a number of leading business people, academicians, and researchers who express concern that educators are producing a workforce that is limited by its inability to retrieve relevant information, critically assess its value and authority, and use it legally and ethically. Ilene Rockman (2004), an information literacy advocate from the California State University's Office of the

---

[2] Approximately one million terabytes equals an exabyte.

Chancellor, states that situations such as this "can negatively affect the ability of businesses to compete and grow in a global economy, and of governments to prosper with increased revenues from businesses and corporations" (p. 9). Information competence is needed to be viable in a knowledge-based economy, yet studies report that students are entering colleges and universities without basic research and information skills (Electronic Publishing Initiative at Columbia [EPIC], 2004; Pew Internet and American Life Project, 2005).

One of the most recognized names in Internet use studies, the Pew Internet and American Life Project (2005), recently found that less than 20% of Internet search engine users can tell the difference between unbiased search results and sponsored advertisements, although 92% of the study participants claim they are confident in their searching abilities. A Columbia University study that examined the impact of electronic resources on scholarly information communication reported that students felt they were "being overloaded with information, and being overwhelmed by the overload of information" (EPIC, 2004). The executive summary went on to state that students expressed difficulty with discerning reliable from unreliable information and that they felt instruction on how to evaluate information would be beneficial.

At a recent Academic Impressions conference, UCF Vice-provost Joel Hartman repeatedly referred to college students' unheralded access to information and its subsequent impact on the traditional academic information model (Hagner & Hartman, 2004). In the past, faculty referred students to the library for their scholarly information needs, where individuals could depend on what were generally considered reputable resources. Now, with the proliferation of the amount of information, and the Internet as

both publishing venue and facilitator to access, facts and data are available virtually everywhere, giving rise to the concern that users may be unaware of the quality of the information they are retrieving (Grassian & Kaplowitz, 2001; Hagner & Hartman, 2004).

Student reliance on the Internet to meet their information demands underscores the value in the ability to evaluate information critically and an understanding of how information is produced and disseminated. Neil Postman (2004) summarizes the state of information as commodity when he writes, "Information comes indiscriminately, directed at no one in particular, in enormous volume, at high speeds, severed from import and meaning" (p. 4). Postman argues that the problem is not lack of information, but the loss of meaning that comes from a glut of unedited, unquestioned, uncombined facts.

Echoing Postman's sentiment, American Library Association (ALA) 2000-2001 President Nancy Kranich (2000) asserts, "More information will not in itself create a more informed citizenry unless people know how to use information effectively to solve problems" (p. 7). As a result, the demand on information users has increased, and not coincidentally the discussion regarding information literacy – the ability to identify, locate, evaluate, and use information effectively – is escalating.

Statement of the Problem

The purpose of scholarly inquiry is to expand, refine, or refute our conceptual or theoretical understanding of phenomena (Postman, 2004). Corollary to this endeavor is the idea that these undertakings will subsequently appear in the literature, thus providing practitioners a means to inform their professional decisions. This, however, appears to be

an unfounded assumption. A number of researchers have reported cognitive or conceptual discrepancy regarding scholarly information access and use. These researchers suggest that students tend to overstate their searching abilities (Fox & Weston, 1993; Greer, Weston, & Alm, 1991; Maughan, 2001), are not consistently critical in their use of information for scholarly argument (Beile, Boote, & Killingsworth, 2003), or feel insufficiently prepared to successfully negotiate the information environment (EPIC, 2004; Kunkel, Weaver, & Cook, 1996; Zaporozhetz, 1987).

Information literacy has recently been recognized by educators and business professionals alike as fundamental to success in a rapidly changing, technology-intensive and information-rich environment. Mary Kennedy (1997) argues that few teachers use the scholarly literature to inform their professional practice because they do not perceive the connection between research and practice. Kennedy suggests that initiatives such as ERIC have been successful in facilitating physical access to the professional literature, but acknowledges that conceptual barriers still exist.

Many students simply do not have the experience or skills to use information technologies effectively. Perhaps this is why information literacy has been recently recognized by a number of professional associations (American Association of School Librarians [AASL] & Association for Educational Communications and Technology [AECT], 1998; International Society for Technology in Education [ISTE], 2000; National Council for Accreditation of Teacher Education [NCATE], 2002) and regional accrediting bodies (cf., Middle States Commission on Higher Education [MSCHE], 2002; New England Association of Schools and Colleges, 2001 [NEASC]; Northwest

Commission on Colleges and Universities [NWCCU], 2003) as fundamental to success in today's information intensive environment.

Additionally, widespread recognition exists for the importance of information literacy instruction as a cumulative and continuous process woven through the curriculum, where skills are developed incrementally through a program of course-integrated instruction (cf., Grassian & Kaplowitz, 2001; Hagner & Hartman, 2004; ISTE, 2000; MSCHE, 2002). Faculty participation and commitment are essential to effective information literacy instruction programs. Kunkel, Weaver, and Cook (1996) report that appropriate assignments are a critical component of information literacy skill acquisition. Faculty must therefore communicate the importance of the research process, allot time for library instruction, monitor student progress, and hold them accountable for the quality of their work. The underlying implication is that the integration of information literacy instruction is the responsibility of all in academia.

Concurrent to these developments is the approval of the Information Literacy Competency Standards for Higher Education (Association of College and Research Libraries [ACRL], 2000), that have the potential to clarify the desired outcomes of information literacy instruction and to provide the possibility for unified assessment efforts (O'Connor, Radcliff, & Gedeon, 2002). Although accreditation standards assign responsibility for information literacy instruction to program faculty, the library's ability to customize information literacy instruction to individual programs places it central to delivery of information literacy instruction in the academy.

Rationale for the Study

Meaningful data from large-scale student assessments is foundational to understanding the breadth and gaps in information literacy instruction (Grassian & Kaplowitz, 2001). Grassian and Kaplowitz offer two reasons for instructional assessment; the first consists of understanding the learning experience in order to improve the experience, and the second is the need to provide program and instructional effectiveness outcome measures for accountability purposes. Assessment data can help determine if education students possess adequate information literacy skills and contribute to the evaluation and revision of institutional information literacy instruction programs. Additionally, in an era of increasing competition for finite resources, assessment data can be used to make the case for the added value of information literacy instruction programs, and for maintaining or increasing their institutional support. For many, an even more imperative reason for assessment exists; assessment offers information about institutional performance for accreditation reviews.

Researchers have developed a number of tools for measuring students' cognitive or affective changes after library instruction, however, the majority of these instruments have been developed for local use only and have not been submitted to rigorous scrutiny. Recognizing the need for an information literacy assessment protocol that measures skills at the undergraduate level, IMLS provided funding for Project SAILS. Project SAILS investigators (2001) cited the need libraries have for outcomes-based assessment that measures the impact of their instructional programs that is valid and reliable. The primary goal of the Project SAILS initiative is to create a test bank of information

literacy items based on national standards that can be used to assess cohort performance with respect to benchmark institutions.

Underpinning this assessment effort is Project SAILS reliance on the ACRL (2000) information literacy competency standards, and instructional objectives developed to support those standards. Much of the utility of the ACRL standards lies in the objectives, which translate the information literacy goals into observable behavior. It may be noteworthy that, while standards relay what students need to know, understand, and be able to do, objectives are more concrete. Objectives describe in empirical terms what students should be able to do and produce, and the characteristics they should possess, upon completion of instruction (Bloom, Madaus, and Hastings, 1981). At the time of this writing, Project SAILS has an ever-growing test bank of approximately 250 items (of which 45 are typically used in any one administration) designed to measure students' cognitive knowledge of related ACRL information literacy objectives. Now in its third year, 80 institutions have participated in the Project SAILS assessment and have received normative results (Radcliff, 2005).

Despite Project SAILS efforts, no current, rigorously reviewed assessment instruments exist that measure information literacy levels of education students. Based on the rationale that "information literacy manifests itself in the specific understanding of the knowledge creation, scholarly activity, and publication processes found in those disciplines" (ACRL, 2000, p. 6), there is an explicit need for assessment instruments that measure information skills unique to the academic discipline. This prompted Project SAILS administrators to offer fellowships for the development of discipline-specific information literacy test items. Procedures followed by the UCF project team to fulfill

requirements of the fellowship are described in Phase I of the study, while Phase II details subsequent procedures for developing and validating a single form assessment instrument.

Goal of the Study

A number of similarities exist between the Project SAILS initiative and the current study. Among them are developing and validating an objective assessment instrument that measures knowledge of information literacy skills levels, provides usability across unique institutional settings, and assesses education-specific information literacy skills. Given this, the question arises as to why an additional information literacy assessment tool for education is needed.

First, Project SAILS principal investigator Carolyn Radcliff (2005) recently announced that discipline-specific test items were in the process of being field-tested, but that further efforts to integrate the items into the general test bank were indefinitely delayed due to redesign of the original general education-level test. Second, intended use of the scales differs, and this makes them distinct from each other in fundamental ways. Project SAILS objectives are to develop discipline specific test items that can be used for large scale administration, with test results being compared to similar institutions.

Where the current study differs from the Project SAILS test is that the B-TILED offers a single form that is relatively inexpensive to administer and score, and that can be used to evaluate instructional efforts at the local level. Therefore, the goals of this study were two-fold: the first was to create a bank of education-specific test items to fulfill

obligations for the fellowship, and the second, to develop and validate an objective assessment instrument that efficiently and inexpensively measures education students' cognitive knowledge of information literacy skills and concepts.

Significance of the Study

This study has potential to be both practically and theoretically significant. First, as the scale is institutionally objective, considerable scope exists to make use of it in replicating information literacy instruction assessment across unique institutional settings. Test results can be used to identify progress of individual students or interpreted as cohort scores, thus offering useful data for providing an outcomes-based quantitative measure for institutional or accreditation purposes. Results also can be used to inform curricular and programmatic decisions, as well as provide for internal and external benchmarking of education students' information literacy skills levels. Assessment can drive integration of information literacy instruction throughout the academic program, and so it may also result in practitioners who use the professional literature to inform their practice and are better prepared to teach and model information literacy skills to their students.

Second, this study has potential to contribute to development of theory in the area. The study was one of the first to meet the criteria set forth by ACRL (2000) to fill the lacunae of discipline-specific information literacy assessment. As such, its development methods may constitute a model for future initiatives. Further, as the test is used, results may help build or verify the theoretical model of information literacy. Information

literacy is a nascent construct and assessment data can help determine what is unique to the theory and define its boundaries and overlap with other constructs. Results of this study will be of interest to faculty involved in information literacy instructional programs, professors of education students, college and university administrators, and program review personnel.

Conceptual Framework

The conceptual framework that guides this study draws from a number of areas. In particular, elements germane to this investigation are contained in the following works: a) the idea of information literacy as an emergent construct (ACRL, 2000), b) information literacy assessment, both in general (Educational Testing Service [ETS], 2004; Project SAILS, 2001) and within the disciplines (ACRL, 2000), and c) development of assessment instruments for library and information literacy skills in education (Morner, 1993).

Among the issues frequently discussed in regard to information literacy is its relation to information technology and information fluency. In this context, information literacy is considered a distinct and broader area of competence than information technology, however, it serves as only one of three components that comprise information fluency. In the following paragraphs, information literacy will first be distinguished from information technology, and then information fluency.

Information technology skills are an integral part of information literacy (ACRL, 2000). A National Research Council (1999) report further distinguishes between

information literacy and computer literacy. According to the report, "computer literacy" is concerned with rote learning of specific hardware and software applications, while "fluency with technology" focuses on a deep understanding of technology and its increasingly skilled use. The ACRL standards add that while information technology "fluency" may require more intellectual skills than learning to use hardware and software associated with "computer literacy," the focus is still on the technology. The differences are further illustrated in the following statement taken from the ACRL standards:

> Information literacy, on the other hand, is an intellectual framework for understanding, finding, evaluating, and using information – activities which may be accomplished in part by fluency with information technology, in part by sound investigative methods, but most important, through critical discernment and reasoning. Information literacy initiates, sustains, and extends lifelong learning through abilities which may use technologies but are ultimately independent of them. (p. 3-4)

Shapiro and Hughes (1996) go so far as to suggest information literacy be conceived of as a new liberal art. They explain that knowing how to use computers and access information is fundamental to information literacy, but it is the nature of critical reflection on the information itself, the technical infrastructure of information, and even its' social, cultural, and philosophical context and impact that is essential to the education of the information-age citizen. These definitions are in contrast to the ISTE National Educational Technology Standards for Teachers (NETS*T) Foundation Standards (2000), that emphasize fluency in information technology rather than information literacy in educational settings. According to the NETS*T standards, fluency with technology is viewed as a goal unto itself rather than as a tool to support access and presentation of information.

The distinction between information literacy and information fluency is less evident. When considering fluency in the context of language or visual design, the fluent individual is generally defined as someone who has integrated the skill to the point that he or she no longer has to consciously think about it, and where performance of the skill is transparent. The Associated Colleges of the South's ([ACS], 2003) definition of information fluency involves the use of critical thinking skills and appropriate technologies to "collect the information necessary to consider a problem or issue, employ critical thinking skills in the evaluation and analysis of the information and its sources, and formulate logical conclusions and present those conclusions in an appropriate and effective way." Therefore, information fluency may be envisioned as the nexus of information literacy, computer literacy, and critical thinking (Rettig & Hagen, 2003), with information literacy serving as one component of the tripartite information fluency model.

The ACS (2003) offers a model of information fluency that is both broader than information literacy and subsumes it as an area of competence, while the ISTE NETS*T focuses on information technology at the expense of print tools and sources. These approaches stress skills that are not considered central to the ACRL (2000) definition of information literacy. With its emphasis on the critical thinking and problem solving skills needed to operate effectively in an information intensive environment, and the use of technology as a tool to facilitate these skills, the ACRL definition of information literacy is used to guide this study, with one qualification. The distinction between information fluency and information literacy lies in differences between the models more

than the definitions of the words. In describing an information literate individual, one of the desired characteristics is to integrate the skill to the point of fluency.

Further, there are seemingly few papers that discuss standardized assessment of information literacy, and even fewer addressing the development and validation of discipline-specific assessment instruments. Researchers have developed tools for measuring the impact of library or information literacy instruction on students' cognitive or affective abilities, but the majority of studies fail to address how the instruments were developed and what procedures were followed to ensure credibility. However, notable exceptions exist and are discussed in greater detail in Chapter Two.

Despite the call by ACRL (2000) for tools that assess specific disciplinary information literacy knowledge, at present no current standardized assessment instruments exist for content areas. With fellowships for development in the areas of biology, communications, education, and history, Project SAILS has been the leader of discipline-specific information literacy assessment. However, with the intention of creating test item banks for large-scale administration, the efforts of Project SAILS are accessible only at the institutional level.

Although no other current information literacy tests for education exist, an early instrument designed to assess library research skill levels of doctoral students in education was discovered. A dissertation project, the Morner Test of Library Research Skills (Morner, 1993) offers an example of an objective assessment scale for library research skills specific to the discipline of education. Created prior to the ACRL information literacy standards, the Morner test relies on skills perceived by experts in the field as essential for successful library use. Recognizing that the instrument is in need of

revision to accommodate technological innovations, Dr. Claudia Morner graciously granted permission for the current study to draw upon her previous work.

Assumptions

Several assumptions underlie this study. Foremost among them is that information literacy skills are an important component of a student's educational experience. Specifically, the author assumes that greater knowledge of information literacy skills and concepts contribute to more scholarly students, more productive workers, and a more informed citizenry. The author also assumes that scores on the B-TILED are indicative of actual skill levels of students. This issue is further addressed in the Limitations section of this chapter and in the Methodology and Results chapters. Finally, there is an assumption that as skill levels increase, performance levels will increase concomitantly and will be exhibited in actual behavior. This assumption, however, may be unfounded. For example, knowledge of ethical use of information does not ensure ethical behavior.

Scope of the Study

This study is characterized by its scope and its limitations. The scope of the study is presented in Delimitations, and defines the boundaries of the study, or what it includes and what it does not. The Limitations section discusses intentional decisions and unconsidered elements of the study that would have made its results more tenable.

Delimitations

Investigation of information literacy curriculum, pedagogy, and instructional best practices is warranted, however, the scope of the current study is limited to assessment of information literacy skills. For the interested reader, a number of works that address information literacy instruction and curriculum integration can be found in the literature.[3] Similarly, evaluation of instructional programs often considers their impact on the affective domain. Studies that assess student attitudes, self-efficacy beliefs, or anxiety levels toward the library and its use were identified, and a section of the literature review was devoted to the topic. However, this study exclusively addresses measurement of information literacy levels in the cognitive domain.

Finally, the scope of this study is limited to developing an instrument that focuses on scholarly information sources relevant to education students. Other instruments designed to assess knowledge of information literacy skills and concepts do exist. Although dated, the Morner Test of Library Research Skills (1993) was developed to measure knowledge of library research skills of doctoral students in education. More

---

[3] The ACRL site links to a wealth of information literacy instruction information. http://www.ala.org/ala/acrl/acrlissues/acrlinfolit/infolitresources/infolitinaction/ilworkshopideas.htm. Other highly useful texts abound. Space precludes listing them all, but two exceptional texts are mentioned here. *Integrating information literacy into the higher education curriculum: Practical models for transformation*, a 2004 book published in San Francisco, CA, by Jossey- Bass and edited by Ilene Rockman, is an excellent text that offers practical examples for integrating information literacy throughout the academic curriculum. For information literacy instruction specific to education, the following text is highly recommended: Shinew, D. M., & Walter, S. (Eds.). (2003). *Information literacy instruction for educators: Professional knowledge for an information age*. Binghamton, NY: The Haworth Press.

recent is the Project SAILS (2001) information literacy assessment instrument and James Madison University's (2004) Information Literacy Test, both developed to assess general education information literacy skills levels. Another instrument is the ETS (2004) ICT (Information and Communication Technology) Literacy Assessment, which appears to be less library-centric and purports to measure a wide range of information and communication technology skills.

Limitations

Methodological limitations exist in that scores on the information literacy scale may not truly indicate education students' skills regarding their ability to locate, evaluate, and ethically use scholarly information. It is generally thought that lower-order thinking skills are more easily measured by objective measures, such as the multiple-choice platform offered by this scale, as opposed to higher-order thinking skills, which are characterized by the critical thinking and problem-solving nature of information literacy.

Further, a single objective measure cannot fully capture students' cognitive abilities or their broader information seeking experiences. Information literacy tests may measure how well students have learned information use skills, but Maki (2002) cautions "they do not demonstrate how well students can solve problems using that information" (p. 10). Measurement of information literacy objectives and associated behaviors may require a more authentic assessment environment. Prolonged interviews and observation of participants' actual information seeking behavior, or rubrics developed to assess the

quality of citations or papers, may be indicated to more fully understand the effectiveness of an information literacy instruction program.

The justification for the selected-response format of the B-TILED is the need for a method that is easy to administer and produces readily analyzable data, the qualification is that multiple forms of assessment are needed to truly gauge student performance and program effectiveness. By definition information literacy applies to the use and evaluation of both library and non-library information sources. This measure attempts to encompass a wide range of information seeking knowledge, but is heavily weighted toward academic information resources.

Definition of Terms

* *Bibliographic instruction*: Instructional programs designed to teach library users how to locate the information they need. Bibliographic instruction usually covers the library's system of organizing materials, the structure of the literature of the field, research methodologies appropriate to the discipline, and specific resources and finding tools. Definition adapted from the *Dictionary for Library and Information Science* (Reitz, 2004).

* *Criterion-referenced tests:* A test where test performance (or score) is interpreted in relation to attainment of specific content or skills. With criterion-referenced testing**, i**t is possible that all, or none, of the test-takers will reach a particular goal or standard.

* *Information literacy*: An intellectual framework for understanding, finding, evaluating, and using information (ACRL, 2000). Grassian and Kaplowitz (2001) add

that transferability, critical thinking, and the ability to apply the information to the individual's life are other important aspects of information literacy.

* *Norm-referenced tests:* A test where test performance (or score) is interpreted in relation to the performance of others who have taken the same test. Tests are often developed so that results fall somewhat along a normal curve, with most students scoring near the middle and fewer scoring low or high.

Parameters of the Review of Literature

In an article that offers guidelines for rigorous test development, Clark and Watson (1995) emphasize the importance of a comprehensive literature review. They note that such a literature review can serve to clarify the nature and range of the target construct, identify problems with existing tests, and indicate whether the proposed scale is actually needed. To this end, an extensive search for information from a number of academic databases was conducted. Databases searched included, but were not limited to, Ebscohost's Professional Development Collection and ERIC, Wilson's Education Fulltext and Library Literature Fulltext, Proquest's Dissertation Abstracts, and the library card catalog and other monographic indexes. Search terms and descriptors were used to find information on the three broad aspects of the study: information literacy and library instruction, assessment of information literacy and library instruction and accreditation-driven assessment, and test construction theory and test development methodologies.

Discrepancies uncovered in studies reported in the current literature led the researcher to consult primary texts and earlier authors on the subject. Conversely, much

of the information on information literacy assessment initiatives is too current to have appeared in the literature, in turn requiring reliance on Web sites, conference presentations, and personal communication. Throughout the process bibliographies from relevant books, articles, and documents were reviewed for additional sources and particular works or authors recommended by professionals in the field were perused.

Summary and Overview of Remaining Chapters

The purpose of this study was to develop and validate an assessment instrument that measures information literacy levels, is tailored to education information sources and examples, can be easily administered and scored, and can be used for both practical and theoretical purposes. The specific approach is detailed in the following sections, but, in general, items were developed based on existing information literacy standards and objectives, reviewed by content experts in the field, submitted to individual student review and small group testing, continuously revised, and then field tested with a sample of education students. Procedures for enhancing the test's reliability and validity were woven throughout its development.

In Chapter Two, the reported research on the state of information literacy instruction assessment is reviewed and discussion in the field regarding test development procedures is summarized. In Chapter Three, the study's Methods are described and a delineation between Part I and Part II of the study is offered. Chapter Four reports descriptive statistics and results of various validity and reliability procedures and analyses. In Chapter Five, concepts covered in the earlier chapters are revisited, practical

and theoretical application of the instrument is suggested, and further directions for research are offered.

CHAPTER TWO
REVIEW OF THE LITERATURE

This chapter presents a summary of the literature related to library and information literacy instruction history and assessment as well as issues pertaining to test validation procedures. The first section documents the history of library instruction and its evolution to information literacy instruction. The following sections contain a discussion of what purposes assessment results can be used and a report of the research on library and information literacy instruction assessment. The research is thematically organized by studies that investigate the impact of instruction on the affective and cognitive domains, and gaps in the literature are identified. Recent assessment initiatives and their current stages of development are also described. The chapter concludes with a discussion of test development issues.

Library Instruction Evolution

Library instruction is not a recent phenomenon. At the first American Library Association (ALA) conference in 1876, Melvil Dewey announced that the library is a school, and the librarian is, in the highest sense, a teacher (Grassian & Kaplowitz, 2001). Library instruction has traditionally been defined as teaching patrons how to use the card catalog and periodical indexes to identify and retrieve information from the library (Salony, 1995). Following this, instruction in the use of libraries remained fairly stable until the advent of the modern library instruction movement, which began in the 1960s. Library instruction came to the forefront with renewed interest in libraries and a

financially supportive climate (Salony). A report funded by the National Commission on Libraries and Information Science (Zurowski, 1974) documents the shift from teaching library instruction that emphasized the acquisition of mechanical searching skills to a more conceptual approach of information use. The author of the report, Paul Zurowski, coined the term "information literacy" to describe this phenomenon. As a results of the report, school libraries, quickly followed by academic libraries, worked to redefine "library instruction" to "information literacy instruction" (Lubans, 1983).

Zurowski's definition was further refined by Patricia Breivik (1985), who describes information literacy as an integrated set of skills and the knowledge of tools and resources, and which includes a critical, evaluative view of the material found. Hannelore Rader (1991) later added that the information literate individual is someone who can be successful in a rapidly changing information environment. Information literacy standards developed by AASL and AECT (1998) reflect this more expanded view. These associations define the information literate student as one who accesses information efficiently and effectively, critically evaluates the information, and uses it accurately and creatively.

While huge strides were being made at the P-12 level, the ALA Presidential Committee on Information Literacy broadened the scope of library instruction by applying the term "information literacy" to information and resources from all venues (Breivik, 1991; Grassian & Kaplowitz, 2001). No longer limiting the term to library resources, information literacy in its current state emphasizes teaching concepts that challenge the learner to engage in problem solving, transferability of searching concepts, and higher order thinking.

The growing concern that students be able to think critically and demonstrate problem-solving behavior, along with the promise of electronic delivery of instruction, has led to a shift in both how and what is taught by instruction librarians (Bober, Poulin, & Vileno, 1995; Chadley & Gavryck, 1989; Salony, 1995). Due to an ever-increasing and complex body of knowledge, a highly dynamic information environment, and need for the continued ability to effectively operate in an atmosphere that will look very different from the present, the traditional goal of teaching basic library skills is no longer sufficient. Instruction in the use of information is now expected to teach the more conceptual approach of information literacy, which engages higher order thinking in the use of appropriate information tools and application of information (Bober et al; Salony).

An instructional framework that extends the teaching of library-specific information resources to a broad array of information tools and resources holds information literacy distinct from library instruction. Therefore, for purposes of this paper, the term "library instruction" will be used to specify an instructional framework that teaches mechanical information retrieval skills, such as how to access and use a particular resource, while "information literacy" will be used to indicate a cognitive strategy that teaches how to select resources appropriate for a topic and the underlying principles useful for effective searching, evaluation, and use.

Assessment of Instruction

In academia, the library's purpose is to support quality educational programs. As such, it is responsible for upholding its part of the academic pact by instructing students

to effectively use library and information sources as an integral part of their learning process. Corollary to this endeavor is assessment, which reveals the level of effectiveness in support of the educational program (Grassian & Kaplowitz, 2001). Assessment of educational quality, often expressed as student learning outcomes, seeks to understand students' skills and behaviors by asking, "What do students know now they did not know before?" and "What can they do they could not do before?" Assessment supports institutional integrity by providing evidence that accomplishments match intentions (Hernon & Dugan, 2004; Rabine & Cardwell, 2000).

Assessment results can be used both to inform internal decisions and to document program effectiveness for external purposes. Understanding how instruction impacts information literacy skills levels is a necessary first step to improving planning, curriculum, and instruction and developing a theory-connected practice of effective instructional techniques (Hernon & Dugan, 2004). Assessment outcomes can also be used to justify an instructional program to library or university administrators and to provide evidence for additional support or resource allocation.

However, assessment results are not limited to revealing how effective programs are for the purpose of internal decision-making they also offer evidence regarding institutional performance for professional and regional program reviews. Information literacy has recently been identified as a key outcome by professional associations and regional accrediting bodies and libraries are now challenged to provide evidence that students effectively use library and information resources as an integral part of the learning process.

The literature on the subject abounds with examples of what constitutes "good" instruction, yet published reports assessing instructional efficacy are limited. Understandably, information literacy is a relatively new construct and rigorous assessment methods are only now forthcoming. Regrettably, if what is published in the literature is any indication of the current situation, many institutions either are apparently not evaluating their instructional programs or are relying on homegrown instruments. These instruments may not be considered acceptable evidence for university administrators and professional bodies that require trustworthy reporting based on credible assessment methods. The following summary of reported research reflects the state of library and information literacy assessment to this point.

Reported Research

Edwards (1994) has reported a three-fold increase in the number of library and information literacy instruction assessment articles published from 1977 through 1991; however, an annual review of research (Rader, 2000) reveals that a considerable number of these publications are program descriptions. Thomas Eadie (1992) states that evaluation studies tend to report on student perceptions or "user satisfaction" of instruction and/or information resources rather than learning outcomes. Surveys from the 1970s and 1980s confirm that evaluation historically has not been a major component of library instruction (Bober et al., 1995; Chadley & Gavryck, 1989).

The literature indicates several barriers to formal instruction evaluation. Patterson and Howell (1990) have observed that most library schools do not offer classes on

instructional assessment, which leaves many librarians to feel they are ill-prepared to properly conduct assessment studies. Formal evaluation may be viewed as too complex or too time consuming, and institutional support may be lacking (Eadie, 1992). Eadie adds that often evaluation is perceived as one more responsibility on an already excessive workload. In addition, library and information literacy instructors may be unwilling to include assessment in their class sessions because it reduces the amount of material that can be included in the limited class time available to them (Grassian & Kaplowitz, 2001). Yet another obstacle to assessment is that librarians generally do not have steady contact with an identified body of students as they matriculate through their programs, thus providing only hit or miss contact with any given student (Rabine & Cardwell, 2000).

Despite these fairly considerable barriers to assessment, a groundswell of instruction evaluation was observed in the 1990s. Bober et al. (1995) surveyed academic library instruction programs, and found that various elements of instruction programs were being evaluated. Program components more often being evaluated include appropriateness and quality of content, methodology used, effect on student attitudes, and impact on student learning. Of these elements, post-instruction attitudinal and learning changes comprise the two outcomes-based assessments reported most frequently. More broadly, these studies have focused on investigating instructional impact on the affective and cognitive domains.

Affective domain. Although evidence of student learning outcomes constitutes primary evidence for assessing educational quality for program reviews, the role the affective domain plays in cognitive learning of library and information literacy concepts has been

recognized by several researchers (Beile & Boote, 2002; Kuhlthau, 1993; Martin, 1989; Nahl-Jakobovits & Jakobovits, 1993; Ren, 2000). For that purpose, a summary of the research reporting on the effect of library and information literacy instruction on the affective domain is included.

For many students, visiting the library and using its resources produce as much anxiety as completing a final exam or giving a class presentation. In a study where students were asked to write down their feelings when talking to reference librarians, students reported feeling confused, worried, and intimidated (Nahl-Jakobovits & Jakobovits, 1993). Researchers have also found that students actually fear talking to librarians, are intimidated by the complexity of search tools, and have low self-confidence in their ability to find information (Mensching, 1987; Nahl-Jakobovits & Jakobovits).

Self-efficacy, a component of Albert Bandura's (1977) social cognitive learning theory, is generally defined as the belief in one's ability to successfully perform a given behavior. While self-esteem is considered a global trait, self-efficacy is associated with specific tasks. A person might enjoy high self-esteem overall, but have low self-efficacy regarding a specific ability, such as effectively using library and other information resources. Partly on the bases of self-percepts of efficacy, people choose what to do, the amount of effort to invest in activities, and how long to persevere at particular tasks. Whether a person makes an effort to handle a given situation depends on the strength of his or her effectiveness beliefs.

Two recent studies have reported significant increases in participants' scores on a measure of self-efficacy between pre-instruction and post-instruction test administration. Wen-Hua Ren (2000) surveyed 85 students before and after instruction as to their self-

rated ability to access and use information in an electronic format. Beile and Boote (2002) likewise surveyed 49 graduate students regarding their self-rated ability to search library databases and retrieve needed information. Post-instruction self-efficacy scores were also compared to grades on an information-seeking assignment and scores on a library skills quiz. In both studies, self-efficacy levels were significantly correlated with cognitive and performance tests; as self-efficacy scores increased, so did knowledge and ability.

In a similar study, Fox and Weston (1993) compared students receiving course-integrated library instruction with students who did not receive formal instruction. Library users self-assessed their ability to successfully use library resources in addition to completing a test to measure actual skills. Based on analysis of survey data, researchers reported students who participated in course-integrated library instruction had higher self-awareness and self-confidence levels than those who did not participate. However, these results did not necessarily translate into actual gains in cognitive skills.

Other researchers have failed to find a significant relationship between self-efficacy scores and actual learning gains. An instructional assessment study of graduating seniors undertaken by the University of California – Berkeley revealed those surveyed held a higher opinion of their library research skills than they were able to demonstrate by their test scores (Maughan, 2001). Greer, Weston, and Alm (1991) also report that while self-assessed library skills were markedly higher for seniors than for freshmen, there was no dramatic trend of increased proficiency from freshmen to seniors in scores on a measure of library skills. Tierno and Lee (1983) explored the impact of course-integrated library instruction on student attitudes and learning. A pre-test was

administered to a sample class on its first meeting and a post-test on its last class. Data, based on the test and interviews, indicated that attitudes were significantly improved toward the library, but learning outcomes were inconclusive.

Researchers (Kuhlthau, 1993; Martin, 1989; Nahl-Jakobovits & Jakobovits, 1993; Ren, 2000) have suggested integrating the affective domain into instructional design, yet little is still known about the impact of instruction on self-efficacy and information literacy proficiency. It is interesting to note that, although every study mentioned use of an instrument, the tool was never offered for review and reliability or validity results were never reported.

Cognitive domain. Researchers have also investigated the impact of library and information literacy instruction on learning outcomes, with similar results. Franklin and Toifel (1994) administered a pre-test to seven classes of education students prior to the beginning of the semester. After one to three hours of library instruction, completion of a library exercise, and a research paper, post-tests were administered. Post-test scores were significantly higher for both graduate and undergraduate students. Bren, Hilleman, and Topp (1998) hypothesized that the most effective method for teaching students the skills and concepts they need to conduct a review of the literature was to emphasize understanding the process rather than the mechanics of searching. The researchers administered a post-test the day following instruction. Students who were exposed to the conceptual learning approach performed significantly better on the test than students who were taught by lecture and demonstration.

Other studies have also reported positive instructional impact on student learning outcomes. Daugherty and Carter (1997) administered a test prior to instruction and a post-test the last class period to investigate whether library instruction participants exhibited greater skill development and improved efficiency compared to non-participating students. As hypothesized, participants exhibited greater skill development compared to non-participating students. Schuck (1992) offered a monetary incentive to attract students to take a library skills test. The significance of library instruction, frequency of library use, and grade point average were used to predict performance on a library skills test. Students who participated in some form of library instruction had a significantly higher mean than the students who did not.

Bibliographic analysis, or bibliometrics, constitutes another popular method for assessing instructional impact. Bibliometrics, originally developed by Kohl and Wilson (1986), is based on scoring bibliographies on the criteria of appropriate type of source for the topic, use of current versus retrospective sources as indicated by the topic, and quality of sources for the topic. Because bibliometrics analyzes the quality of the end product of the instruction, proponents believe it is a better indicator of learning than using library skills tests.

In a study designed to assess the impact of conceptual learning as compared to a traditional library skills class, Kohl and Wilson (1986) randomly selected student term papers from two groups of students, one group received library skills instruction taught as a cognitive strategy, and the other group was taught mechanical search skills. Analysis of citations indicated that library instruction, actively taught as a cognitive strategy approach, produced significantly higher quality student bibliographies. Dykeman and

King (as cited in Bober et al., 1995) also analyzed term paper references to assess learning outcomes of library instruction. They found the group receiving instruction produced better written papers that contained more subject pertinent research material, however, they acknowledged it was hard to control for extraneous factors.

Two studies that failed to find that instruction positively impacted student learning outcomes were Ackerson, Howard, and Young (1991) and Cameron (2004). Using bibliometrics as an assessment method, Ackerson, Howard, and Young (1991) compared data on the number of library instruction sessions students received. Replicated over a period of five semesters, classes that received one instructional session were compared to those who received four sessions. There was a statistically significant difference in the scores between the two classes in only one semester. Likewise, Cameron (2004) did not find a statistically significant difference in scores between students who completed a tutorial and those who received no instruction at all.

Summary of reported research. Researchers who reported positive post-instruction statistical significance, whether for affective or cognitive impact, include Beile and Boote (2002), Ren (2000), Bren, Hillemann, and Topp (1998), Daugherty and Carter (1997), Dykeman and King (as cited in Bober et al., 1995), Franklin and Toifel (1994), Schuck (1992), Kohl and Wilson (1986), and Tierno and Lee (1983). Other research (cf., Ackerson et al., 1991; Cameron, 2004; Fox & Weston, 1993; Greer et al., 1991; and Maughan, 2001) has failed to find a statistically significant relationship between instruction and attitudinal or learning gains. Many of these studies reported positive post-instructional impact, while others did not. Results of these studies may have been

interpreted by researchers to have local value, in that some quality of instruction was explained, but when considered in their totality they offer little in the way of enhancing our understanding of information literacy as a theory.

With few exceptions, these studies used locally-produced evaluation tools that had not been subjected to rigorous scrutiny or developed for administration across unique institutional arrangements. When discussing impediments to formal instruction evaluation, none are as problematic as the lack of a global assessment instrument (Barclay, 1993; Bober et al., 1995). Certainly, systematic evaluation of library and information literacy instructional programs that uses credible assessment methods would facilitate both evaluation of instruction for programmatic purposes and contribute to our understanding of information literacy and its viability as a theory. Fortunately, significant progress has been made in the development of credible instruments in the last several years.

Current Initiatives

As information literacy skills are interpreted in the context by which they are applied, project goals and key student learning outcomes should be identified prior to choosing an appropriate assessment method. For example, if a goal of the instructional program is for students to produce better researched and documented papers, then rubrics (see Beile et al., 2003) can be applied to the reference list to evaluate the quality of sources cited by the student. Similarly, if a goal of the project is for students to critically analyze the research in their field and synthesize it in papers, theses, and dissertations,

then tools that evaluate the sophistication of thesis development and structure of the literature review can be employed (see Boote & Beile, 2005). However, analyzing writing samples or term paper references is admittedly labor intensive, and their utility may be limited to selected students or courses.

Instead, not forgetting the allure of documentation for program reviews and much more practical for large-scale assessment, current initiatives have pragmatically focused on development of objective tests. Several tests are receiving widespread recognition; among them are James Madison University's Information Literacy Test, the ETS ICT Literacy Assessment, and the Project SAILS information literacy test. A description of each of these projects and their current state of development follows. Many of these efforts are too recent to have appeared in the scholarly literature. Therefore, much of the information reported here is based on personal communication and collaboration, conference presentations, and information gleaned from Web sites and postings to electronic listservs.

James Madison University's Information Literacy Test. The Information Literacy Test is a computerized, multiple-choice test developed collaboratively by the James Madison University Center for Assessment and Research Studies and the Libraries. It is designed to assess ACRL information literacy standards 1, 2, 3, and 5 at the general education level (James Madison University, 2004). The test does not address Standard 4, as this competency is not easily measured by a multiple choice format test.

Steven Wise (personal communication, August 15, 2005), one of the project directors, notes the test can be used for program assessment or to test individual student

competency. The Information Literacy Testt contains 60 items, plus five pilot items, and takes approximately one hour to complete. It is Web-administered, multimedia-intensive and contains 41 knowledge items and 19 items to assess application of knowledge. The test has been in development a little over two years and is available for administration by other institutions.

The ETS ICT Literacy Assessment. Another instrument which recently underwent extensive field testing is the ICT Literacy Assessment. ETS (2004) and a group of colleges and universities have worked together to create the test, which purports to measure students' cognitive and technical skill levels in an authentic, technology-intensive environment. At the time of this writing, the test has been administered to approximately 4500 students across 31 campuses (Katz, 2005). This test, likewise, is multimedia-intensive and delivered via the Internet. Administration time is approximately 2 ½ hours, and it includes both cognitive problems and attitudinal and demographic questions. It is possible shorter versions of the test, in the 50 to 75 minute range, will soon be available. The test is based on information, communication, and technology standards from ACRL and ISTE. Test results are currently provided at the institutional level, but individual scores may be forthcoming.

Project SAILS information literacy test. The most longstanding of these large-scale assessment initiatives is Project SAILS (2001). A federally funded initiative, Project SAILS evolved in response to the need for an information literacy assessment instrument that can be administered across institutions and that provides data regarding institutional

performance when compared to identified benchmark institutions or ACRL standards. To date, Project SAILS has approximately 250 test items in its bank and has recruited 80 institutions to participate in assessment of their information literacy instruction programs (Radcliff, 2005). Items are randomly selected from the test bank, each test contains 45 items. As with the James Madison University and ETS tests, the Project SAILS test is designed to evaluate information literacy skills at the general education level. These skills are general in that they are not specific to any particular discipline.

Project SAILS administrator Carolyn Radcliff (2005) recently reported that the initiative is at the end of its initial three year grant funding. The project is expected to continue, but developers will be taking a one year hiatus to create a custom interface for participating institutions, enhance administrative tools, translate the test into different languages, and further analyze test data. The test has been administered, either electronically or in print, to over 39,000 students. The next administration of the test is tentatively planned for fall 2006.

Summary of current initiatives. Information literacy assessment has gained broad-based national attention, and is meaningful for at least two reasons. First, this attention emphasizes the importance being placed on information literacy as a key student learning outcome by the academy. Second, it illustrates the need for credible instruments to assess knowledge of the construct. Given the emergent state of information literacy assessment, test development efforts have rightfully focused on assessing general education level information competence, which, in turn leaves discipline-specific assessment a future initiative. However, based on the rationale that "information literacy manifests itself in

the specific understanding of the knowledge creation, scholarly activity, and publication processes found in those disciplines" (ACRL, 2000), the development of assessment instruments unique to the academic discipline remains essential. The current study is among the first of its kind to venture into developing and validating a discipline-specific assessment test.

Test Development Theory

Achievement tests are the oldest and most widely used type of measuring device, with earliest use documented to China and dating back several thousand years (Davis & Diamond, 1974). Originally used to assess applicants' knowledge for civil service positions, achievement tests have served as precursors to student learning examinations that have been administered by schools throughout history. Davis and Diamond add that by the mid-20th century a rich literature with prescribed methods for identifying content domains, procedures for preparing test items, and statistical techniques for interpreting test results had evolved. The 1950 publication of Guliksen's *Theory of Mental Tests,* which articulated the theory of statistical measures applied to achievement test development, led many researchers to conclude contributions to classical test theory were replete.

Based in large part on the influx of immigrants into the United States in the late 19th and early 20th centuries, the main purposes of testing were for placement and advancement decisions (Davis & Diamond, 1974). These tests were generally referred to as norm-referenced tests and they were used to determine individual performance in

relation to others. Norm-referenced measures distribute scores along a bell-shaped curve and are most appropriately used when comparisons across numbers of students need to be made. However, with the advent of more specific and observable objectives writing, criterion-referenced tests emerged at the expense of nationally-normed standardized achievement surveys (Davis & Diamond).

Fundamentally different from norm-referenced tests, Nitko (1970) defines a criterion-referenced test as one that is "deliberately constructed to give scores that tell what kinds of behaviors individuals with those scores can demonstrate" (p. 38). Nitko (1974) further suggests that criterion-referenced tests must include the following four characteristics: classes of behaviors that are specified as clearly as possible before the test is constructed, each behavior class is represented by a set of test items, a sampling plan is used to select the test items, and the obtained score meaningfully and objectively expresses the individual's performance characteristics in these classes of behavior.

Assessment results today are used for a variety of purposes. Skager (1974) suggests that test results can be used for curriculum planning, classroom management, instructional performance, resource allocation, accountability, and prediction of future academic achievement or employment success. Other researchers (cf., Harris, 1974a; Nitko, 1970) add the need to certify student attainment for purposes of assigning a grade or placement within the curriculum. These purposes constitute two broad categories: test results are used to inform decisions about individuals and about treatments.

While norm-referenced measures are used to identify an individual's performance in relation to the performance of others on the same measure, criterion-referenced testing is used to identify an individual's status with respect to an established standard of

performance. When considering the development of norm-referenced and criterion-referenced tests and their purposes to which they are used, it is useful to compare and contrast the two approaches as significant conceptual and methodological differences exist.

Norm-referenced measures are often used when a degree of selectivity is required by the situation and relative comparisons are needed, such as professional or academic entry into a small number of openings. Norm-referenced tests try to tease out variability among test takers. This purpose of norm-referenced testing differs from criterion-referenced measures, which seek to determine whether an individual has mastered an objective or skill. After instruction, a criterion-referenced test score should give some indication of instructional efficacy. With criterion-referenced testing, variability of scores is not central to the test, as it is possible for 100% of students to possess the knowledge or skill being assessed.

Test Development Issues

The purpose to which a test is to be used has significant implications for its validation. With norm-referenced testing there was general acceptance of how tests should be constructed and judged (Davis & Diamond, 1974; Popham & Husek, 1969). However, the emergence of criterion-referenced testing led many researchers to question the logic of applying the same development procedures to tests with different purposes. Although a number of researchers (cf., Harris, 1974a; Popham & Husek, 1969; Simon, 1969) agree that one test can serve many functions, and even that norm-referenced tests

and criterion-referenced tests can be constructed in the same manner, they also acknowledge that interpretation of test results differs as do procedures to validate the instruments.

For criterion-referenced tests, the distribution of student scores who have mastered the instruction or have the construct being measured is heavily skewed, as all students are expected to succeed. A huge range of variability in scores of instructed students is not considered optimal, in that it may indicate instructional deficiencies. This range in variability, of course, is not the case with norm-referenced measures, which seek to have test results distributed in a normal bell-shaped curve. Norm-referenced items often produce great variance among scores, and use of item analysis to select items with positive discrimination indices (items answered correctly by high scoring test takers but incorrectly by low scoring test takers) is of great importance. This is not necessarily the case with criterion-referenced tests.

The following excerpt from Popham and Husek (1969) presents a very cogent and convincing summary of the concept of variability as applied to validation of criterion-referenced and norm-referenced measures.

> The issue of variability is at the core of the difference between norm-referenced and criterion-referenced tests. Since the meaningfulness of a norm-referenced score is basically dependent on the relative position of the score in comparison with other scores, the more variability the better… With criterion-referenced tests, variability is irrelevant. The meaning of the score is not dependent on comparison with other scores, it flows directly from the connection between the items and the criterion. (p. 3)

The consequences of the relevance of variability permeate any discussion of the two approaches to testing. Numerous published procedures and study methodologies

exist that promote procedures that are based on the desirability of variability for instrument validity and reliability. Popham and Husek (1969) state that, as many procedures for assessing validity are based on correlations, and thus on variability, results of these types of procedures are of secondary importance for validating criterion-referenced tests.

A number of errors can occur when procedures for validation of psychological concepts are applied to criterion-referenced tests. Nunnally (as cited in Thompson & Daniel, 1996) offers that factor analytic procedures are at the heart of *psychological* constructs [italics added]. Other researchers, including Davis and Diamond (1974), Horn (1966, 1968), Popham and Husek (1969) and Skager (1974) have substantiated this view. Simon (1969) adds that the use of criterion-referenced tests is appropriate to instruction with specified objectives and whenever mastery of subject matter is of prime concern. He concurs that many statistical measures relying on variability are of secondary importance for validating criterion-referenced tests.

Discussions of test validation principles date back more than 30 years in the literature, yet many test development studies reported in the literature today appear to be unaware of these core principles. In a survey of 41 scale-development articles published from 1989 through 1994, Clark and Watson (1995) found basic test development and validation principles were not being universally honored. They state, "…widespread misunderstanding remains regarding precisely what construct validity is and what establishing construct validity entails" (p. 310), and add that construct validity cannot be inferred from a single set of observations. However, it is not uncommon to find factor analytic results reported as "validity" and correlation coefficients used to attest to an

instrument's "reliability" in the literature. Neither factor analysis and validity nor internal consistency and reliability are interchangeable concepts. Perhaps even more important is the general lack of recognition that these procedures are not the most essential for validating objective, criterion-referenced tests.

It is generally accepted that objective, outcomes-based assessment instruments call for a number of procedures to verify their credibility. The importance of the following procedures for criterion-referenced tests varies, but most test validation studies report findings for content validity, criterion-related validity, factor analysis, and stability and internal consistency procedures. Although Popham and Husek (1969) state that analytic procedures based on variability of scores should not constitute primary evidence for scale validity, they do concede these measures can provide valuable information. Consequently, results from each of the aforementioned procedures are reported for this study.

In summary, studies reported in the current literature appear to indicate that test development procedures are being conducted and reported without an understanding or consideration of appropriate methods. It is for this reason that test development theory advocated by early pioneers in criterion-referenced testing is reviewed. This study uses statistical procedures and reporting and interpreting conventions generally found in scale validation reports, but also presents viewpoints of early researchers regarding the role and importance of those procedures.

Cut Score Calculation

A final test development issue addressed in this paper is the calculation of the cut score. Many library and information literacy instruction evaluators seek a single numerical value that represents to what extent scores on a test reflect proficiency of the concept. More specifically, these evaluators want to know what number or percentage of correct answers should be put forth to indicate mastery. The cut score separates test takers into two groups; those inferred to have attained competency, and those who have not (Berk, 1986; Ebel, 1968; Harris, 1974b; Popham, 1974).

Tests are often developed based on student learning outcomes that indicate the desired level of performance. For example, it is not unusual to find a learning objective and related assessment item with a target that states "All students will answer the item correctly100% of the time." These performance levels, or targets, are often used to indicate a passing score. This scenario, however, is not without problems. Popham (1974) explains that use of target proficiency levels for test scoring is extremely complicated as target proficiencies are assigned at the objective level, yet in reality, a test is composed of any number of items with varying targets. It is highly likely a test could contain an *x* number of items requiring all students answer the items correctly 100% of the time, and a *y* number of items that stipulate correct answers 80% (or some other variation) of the time. For this scenario test administrators must decide how to score tests and interpret their results for indication of proficiency.

The debate over whether individual passing scores should be established for criterion-referenced tests is long-running. Researchers (cf., Lord, 1974; Nitko, 1970;

Simon, 1969; Skager, 1974) have acknowledged it is difficult, if not conceptually flawed, to try to arrive at a unique score from continuous data that is supposed to be indicative of competence. Even proponents of the practice, such as Ronald Berk (1986), who suggested various methods for calculating passing scores, concede that the cut score is an "artificial dichotomy" imposed on continuous test score distribution. Researchers also recognize that it is a practice considered essential for many educational decisions.

In lieu of other guidelines, these "competency thresholds" are often arbitrarily set. For example, in interpreting mastery states of students for an information literacy test, Cameron (2004) decided that, on a 33 item cognitive test of information literacy skills, raw scores of 32 to 33 correct indicated exceptional competence, 25 to 31 were competent, 21 to 25 minimally competent, and 20 and below incompetent. Gratch Lindauer and Brown (2004) offered that, as 70% correct was generally acceptable for students to receive a "C" grade, then the same percentage could be used to indicate competence on an information literacy test. Lord (1974) explains that if two tests are devised that measure the same objective or skill, and they are not equally difficult, then passing scores will differ at varying ability levels. Lord concludes that passing scores cannot be arbitrarily assigned to criterion-referenced tests.

Researchers who calculate cut scores approach the process from two different perspectives; those who view competence as an all or nothing state, and those who see competency as a continuously distributed ability (Berk, 1986). The "state" perspective supports setting standards at 100%, then working back to a smaller percentage to accommodate measurement error. Those in the "continuous" camp argue that, as ability is continuously distributed, then competence can be determined by defining its

boundaries. Berk proposes the perspective can be content-dependent. The state approach is justified for skills that satisfy the assumption of all or nothing mastery of homogeneous content while other, less discrete content areas can be viewed as continuously distributed. An example of content that would qualify for 100% mastery is mathematics, where performance mastery of identified objectives is essential before one can proceed to the task of acquiring higher skills. Certainly, a case can be made that information literacy competence, with a range of heterogeneous content and no hierarchal structure (where it is not the case that some skills need to be learned before acquiring higher ones), could be viewed as continuously distributed.

Berk (1986) suggests recent test developers have overlooked the literature of the 1960s and 1970s, where passing score calculations have received considerable attention. Given that minimal attention has been extended to the literature on cut score calculation by researchers reporting on mastery states for information literacy, a synopsis of Berk's discussion of methods for setting or adjusting cut scores is warranted here. In a review article, Berk summarizes 38 cut score calculation methods and evaluates them based on practical and technical criteria and performance reliability. Berk sorts the methods into three distinct categories: judgmental, judgmental-empirical, and empirical-judgmental. Judgmental methods are based primarily on the judgments of one or more persons without the benefit of any performance data. Judgmental-empirical and empirical-judgmental methods are compromise procedures; they rely on judgments of one or more persons, with performance data available to guide those judgments. The support for judgmental-empirical methods rests primarily on judges' decisions, while empirical-

judgmental methods are based primarily on performance data and statistical analysis of that data.

Berk (1986) advises that the role of judgment should not be underrated when setting the cut score, and suggests an eclectic approach that pulls together various components from the methods reviewed. His recommendations draw from both judgmental and empirical procedures. They include a panel of experts to initially judge items, then present the judges with item difficulty levels to allow them to adjust their decisions in a two-step process. These procedures are culled primarily from the iterative Angoff method and the informed judgment methods, both of which scored highly against Berk's evaluation criteria. Berk cautions that the initial cut score calculation is a preliminary procedure, and that empirical-judgmental methods should be used to adjust the suggested score after administering the test. These procedures are further addressed in the Methodology and Results chapters.

Summary

Only in the last forty years has modern library and information literacy instruction evolved into a standard reference service (Chadley & Gavryck, 1989). As evidenced by the growth in the literature, there is increasing interest in the evaluation of these instructional programs (Bober et al., 1995). The literature indicates, however, that libraries are only now beginning to systematically evaluate their library instruction programs (Barclay, 1993; Chadley & Gavryck, 1989). In this climate of academic accountability where electronic information services are changing the kinds of skills and

techniques necessary to use information resources effectively, the quantity and quality of assessments will have to improve (Bober et al., 1995; Schuck, 1992).

Although a consensus exists among librarians that information literacy instruction affects the scholarly output of students, consistent data relating library instruction to improved works of scholarship is lacking. Systematic evaluation of information literacy instruction is needed for program reviews, to inform instructional and curricular decisions, and to provide a deeper understanding of the construct. Several current, large-scale test development initiatives were described. The B-TILED is yet another tool in the nascent, but burgeoning assessment movement, and is among the first to investigate information literacy as it pertains to a specific discipline.

This chapter also included a summary of the conceptual and methodological differences that exists between norm-referenced testing and criterion-referenced testing. The concept of variability and its role in validating objective instruments was presented and reasons for reviewing the early literature on the topic given. The section ended with a brief discussion of the purpose of determining cut scores and approaches to calculate the value. These themes run throughout the paper and are discussed more fully in the following chapters.

CHAPTER THREE
METHODOLOGY

The goal of the study was to develop and validate an instrument that assesses information literacy skill levels of education students. The test can be used for purposes of examining the impact of information literacy instructional programs and providing evidence to accrediting bodies and professional associations. The study's Methods are comprised of two main sections; Phase I and Phase II. Phase I describes work performed for the Project SAILS fellowship, awarded to develop test content and populate a test item bank for education-related information sources. Phase II of the study explains procedures for the subsequent development of the instrument.

Objective, criterion-referenced assessment test development begins with the process of translating basic elements of a theory into directly observable variables. Each of these theoretical elements must be operationally defined in order to be measurable. Error is often introduced when translating these elements into measurable variables. A number of validity and reliability checks are recommended to ascertain how well the instrument actually measures the construct. A brief description of these procedures follows, and each is expanded upon in the appropriate section.

Validity refers to the accuracy of the scale and seeks to determine how well the instrument measures what it intends to measure. The primary purpose of test validation is to investigate how well the test translates, measures all parameters, and is restricted to only the construct. Many types of validity, such as content validity, criterion-related validity, and factor analysis, are reported in the literature. All are used to provide evidence of construct validity. Content validity, which seeks to ensure test items are an

accurate reflection of the criterion, constitutes primary evidence for construct validity of criterion-referenced tests (Horn, 1966, 1968; Popham & Husek, 1969; Simon, 1969). Another facet of validity is criterion-related validity, which is concerned with how test results compare to an external criterion, such as an already validated test or actual performance. Criterion-related validity asks whether high scores on the test are a good indication the student can execute behaviors associated with high skills levels. Factor analytic procedures are less compelling for criterion-referenced tests, but are offered as evidence of whether the instrument is tapping into the qualities it is designed to measure.

Reliability is generally defined as how accurate a test is in discriminating among knowledge levels. Two procedures used to offer evidence of reliability are internal consistency and stability. Internal consistency refers to the degree to which the items measure the same construct. For purposes of analysis, high internal consistency coefficients indicate questions are related in terms of who answered them correctly, and low coefficients indicate when questions are unrelated in terms of who answered them correctly. This procedure is peripheral to criterion-referenced test validation for two reasons. First, information literacy is a very heterogeneous construct, and items are not expected to be cohesive. Second, the procedure relies on variability among scores. Similar to factor analytic results, internal consistency values may offer evidence that the test is measuring the students' true score rather than measurement error, but low values do not necessarily mean the test is unreliable. A procedure more central to criterion-referenced test reliability is stability, which looks to see if similar test results are attained over repeated test administrations.

Additional analytic procedures discussed in the chapter include item analysis, which is used to assess the quality of items by analyzing difficulty levels, item discrimination indices, and responses to distractors. A final procedure used to establish a preliminary passing score for the test is presented.

Ethical Considerations

To minimize potential harm to participants this study was conducted in accordance with all federal and university mandates, and consonant with professional standards for conducting research. The research proposal was reviewed and approved by the university Institutional Review Board (IRB) and permission to administer the test was obtained from all participants. The IRB approval document and participant volunteer form are presented in Appendix A. This study also adhered to practices recommended by Creswell (1994), who suggests that, if and when ethical dilemmas arise, participants' rights, interests, and wishes are considered first when decisions are made.

Phase I of the Study – Project SAILS-supported

The Project SAILS fellowship guidelines stipulated much of the timeline and procedures of the initial phase of the study. Four project teams, covering the subject areas of biology, communications, education, and history, attended a week-long training institute at Kent State University in April 2004. Upon returning, the UCF project team for education presented a work plan to Project SAILS administrators defining the scope

and nature of the test content and timeline of the project. After receiving authorization to proceed, the UCF project team began to develop test items.

Test Content

The creation of an initial item pool was a crucial stage in constructing the scale. Post-data collection analysis identifies items that can be dropped, but cannot detect content that should be included in the scale (Popham, 1974); no post-data collection analyses can remedy deficiencies in an item pool (Clark & Watson, 1995). Clark and Watson add that the fundamental goal of item writing is to ensure that the initial pool is broader than one's own theoretical view of the construct, and that the pool also includes content that is tangential to the construct. Information literacy is a widely recognized construct that recently has appeared in a number of professional and accrediting associations' outcomes criteria. Most often, these criteria expressed as standards and goals, are prerequisite to focusing instructional efforts and serve as markers for the evaluation process (cf., Elliot, Kratochwill, Littlefield, & Travers, 1996; Linn & Gronlund, 1995; Mager, 1997). Following training, several general and education-specific standards documents were reviewed to identify possible learning objectives.

The most comprehensive standards, in that a number of learning outcomes and objectives have been developed to accompany them, are the ACRL (2000) Information Literacy Competency Standards for Higher Education. The intent of the standards is to provide a framework for determining whether or not a person is information literate from a higher education standpoint. These standards were chosen as the basis for this study

(see Appendix B). However, standards that apply to teacher education accreditation efforts also exist and were examined for applicability.

After further review of standards from NCATE (2002), the Interstate New Teacher Assessment and Support Consortium ([INTASC], 1992), ISTE (2000), and the ACRL Education and Behavioral Sciences Section ([EBSS], 1992), the project team decided to rely upon the ISTE (2000) NETS*T foundation standards. Although NCATE standards address information literacy skills,[4] the ISTE NETS*T are more developed and often cited for accreditation purposes. The ISTE NETS*T standards (see Appendix C) were aligned with the ACRL objectives to form a blueprint for test content development.

Portions of the NETS*T that were identified as relevant to information literacy skills include:

- **Standard II: Planning and Designing Learning Environments and Experiences**
  Section B: Teachers apply current research on teaching and learning with technology when planning learning environments and experiences.
  Section C: Teachers identify and locate technology resources and evaluate them for accuracy and suitability.

- **Standard V: Productivity and Professional Practice**
  Section A: Teachers use technology resources to engage in ongoing professional development and lifelong learning.

- **Standard VI: Social, Ethical, Legal, and Human Issues**
  Section A: Teachers model and teach legal and ethical practice related to technology use.

---

[4] The NCATE 2002 publication *Professional Standards for the Accreditation of Schools, Colleges, and Departments of Education* states "they (teacher candidates) are able to appropriately and effectively integrate technology and information literacy in instruction to support student learning" (p. 19)

Overall, these standards encompass identifying, accessing, locating and using research and technology sources, evaluating sources, and modeling and teaching legal and ethical practices related to technology use. Because the areas were too comprehensive for the scope of this initial project, they were further refined; objectives that targeted conceptual understanding or universal application were emphasized at the expense of objectives that focused on institution-specific arrangements or holdings. The first broad area (identifying, accessing, locating, and using research) was subdivided into two parts; 1) identifying, evaluating, and selecting finding tools and 2) demonstrating knowledge of general search strategies. The team took a more holistic view of the next area (evaluating sources) and extended this standard to include selection as well as evaluation of sources. Finally, as this instrument is not designed to measure actual behavior (i.e., "modeling and teaching" legal and ethical practices), the fourth area is comprised of objectives designed to assess a candidate's understanding of those legal and ethical issues.

The ISTE standards were regrouped into the following content clusters:

- Identifying, evaluating, and selecting finding tools
- Demonstrating knowledge of general search strategies
- Evaluating and selecting sources
- Demonstrating knowledge of legal and ethical practices

A map that illustrates the link between the specific content clusters derived from the NETS*T and ACRL information literacy objectives is offered in Appendix D. This map, and a rationale for coverage of the test content, was submitted to Project SAILS administrators for approval.

Test Item Construction

After Project SAILS staff reviewed and endorsed test content parameters, the UCF project team began writing items designed to measure students' levels of information literacy skills as related to identified objectives. Project SAILS previously determined that the multiple-choice item format was most efficient for assessing a large population. To facilitate test scoring, given that education-specific test items might be used in conjunction with items from the general information literacy test bank, Project SAILS personnel prescribed that project teams also adhere to the multiple-choice item format, with multiple correct answers and varying numbers of response choices allowable.

Project SAILS administrators further suggested that development teams identify 30 to 40 objectives, and then write items to assess cognitive knowledge of those objectives. Ideas for items were culled from a variety of sources, including those borrowed or adapted from the Project SAILS test bank and Claudia Morner's (1993) library skills test. The majority of items, however, were original and written by the development team.

This phase of the project was time consuming and further complicated by the need for institutionally-objective items. One factor that has been recognized as an impediment to development of widely applicable standardized tests is that libraries differ in their physical arrangements and even in the databases to which they subscribe (Barclay, 1993). As test items were supposed to reflect databases and sources unique to the field of education, a major challenge of test item development was to write items that

could apply to any number of unique institutional arrangements without diminishing the value of the test.

Preliminary item writing resulted in a bank of 58 test items. Review by the UCF project team revealed that additional items were needed for the area of ethical use of information. Four more items were written to address this category, bringing the test bank to 62 items. A map of the original items to NETS*T content clusters and ACRL objectives is offered in Appendix E. Test items were sent to Project SAILS and content experts for review in June 2004. The following example illustrates the relationship between the content cluster, the NETS*T standard, and the ACRL objective with the item.

> Content Cluster A
> Identifying, evaluating, and selecting finding tools
>
> NETS*T II.C.
> Section C: Teachers identify and locate technology resources and evaluate them for accuracy and suitability.
>
> ACRL objective
> 1.1.3.2 Demonstrates when it is appropriate to use a general and subject-specific information source (e.g., to provide an overview, to give ideas on terminology).
>
> Item
> You have been assigned to write a paper on the whole language movement, a topic with which you are unfamiliar. Which of the following is the best source to find a brief history and summary?
> CHECK ONLY ONE ANSWER.
> a. a book titled *Perspectives on whole language learning: A case study*
> b. a dissertation titled *Whole language and learning disabilities: Case study of a student teacher's beliefs development*
> c. a recent newspaper article titled "Whole Language in the Classroom: How Effective?"
> d. *Encyclopedia Britannica*
> e. *Encyclopedia of Education*

Further Development

To meet the timeline goals agreed upon in the work plan, it was necessary to continue developing, testing, and revising while items were sent to Project SAILS coordinators and content experts for review. The procedures described in this section were adhered to for the purpose of providing evidence for content validity. One-on-one testing was conducted using a think-aloud protocol, and small group administration of a draft version of the instrument ensued.

Content validity. Content validity concerns itself to the degree which a test is representative of the content it was originally designed to test. Researchers (cf., Horn, 1966, 1968; Popham & Husek, 1969; Simon, 1969) offer that representativeness of content is the primary evidence for construct validity of criterion-referenced tests. Clark and Watson (1995) point out that theory articulates what a construct is, and what it is not. Theory clarifies the nature and range of the content of the target construct and is essential to good item writing. To determine the degree of match between the items and the objectives a panel of experts in the field rated the items on the strength of accuracy, clarity, and institutional objectivity. The five experts are on faculty at the following institutions: Syracuse University, University of Connecticut, University of Kansas, University of New Hampshire, and University of Wisconsin-Madison. All reviewers have library science degrees, work extensively with education students, and have a wealth of experience with both education information sources in the field and information literacy competencies and standards.

After agreeing to serve as content reviewers, the experts were sent a packet with a cover letter explaining the scope of the project, a copy of the test with correct responses, an item relationship map to ACRL objectives and ISTE NETS*T, and a rating grid. The packet was hand delivered at the ALA Annual Conference in Toronto, CN to two reviewers and mailed to the remaining three.

The experts were asked to judge each of the items by assigning a rating of 0 (for *nonexistent*), 1 (for *low*), 2 (for *moderate*), and 3 (for *high*) levels of:

- accuracy - How accurately does the item reflect the ACRL objective?
- clarity - How clearly written and understandable is the item?
- institutional objectivity - Does any of the content of the item reflect local arrangement or can the item be applied across multiple settings?

The following instructions were offered to the reviewers for further clarification. "Using the *accuracy* category, for example, if you think the item is not related to the objective at all, assign a 0. Conversely, if the item is very aligned to the objective, assign a 3. Scores of 1 and 2 can be assigned to moderately related items that fall between the two extremes. The same scoring format should be applied to each of the categories."

To compare results from the five experts, item scores were summarized and placed into a grid. This method was adapted from Morner (1993) and allowed for quick comparison of differences and similarities in experts' ratings. Overall, the raters appeared to be favorable regarding the item rating for the criterion of accuracy. Of the 62 items reviewed, only 3 items (4.8%) received average scores of below 2.0. The mean average for all five experts was 2.0 or better for over 95% of the items. Eight items, or 12.9%, received an average rating of under 2.0 for clarity, and 1 item (1.6%) was below 2.0 for institutional objectivity. Based on Morner's protocol, it was decided in advance

that items scoring below an average of 2.0 on accuracy, clarity, or institutional objectivity would be revised or flagged as potentially problematic for Project SAILS personnel.

One-on-one testing. Individual testing was conducted for several weeks throughout July 2004. A combination of six newly hired and continuing library student assistants answered each item individually, using a think-aloud protocol to articulate their understanding of the item, their choice of answer, and why each of the other choices was eliminated as a possible correct answer. Students were four females and two males, five native English speakers and one with English as a second language, ranging from freshman to senior. Test administrators included the UCF project team and a visiting librarian from the University of West Indies - Mona, Jamaica campus.

Utilization of the think-aloud protocol was a crucial step in construction and development of the test items. Great possibility exists for the potential of researcher bias while developing an instrument. Bradley (1993) suggests that the researcher's preunderstanding, or the conditions that impact the selection of the research question or problem, can affect the conduct of inquiry. As such, clarifying the researcher's own preunderstanding is an important step in the design of the study. Budd (1995) applies Bradley's comments to library and information studies when he contemplates the ontology of the library, and the tensions between the emic and etic perspective of it:

> We cannot forget that the entirety of the library signifies, directly or indirectly, the product of intentionality. The catalog, the physical and conceptual organization, even the physical structure itself are consciously created by an I. The library user – another I – adopts an intentional stance when perceiving the aspects of a library. To the user, then, the library is the other. (p.312)

Addressing concerns about etic constructs and end-user's emic perceptions, Peter Ingwersen (1982) used think-aloud eliciting procedures to tap into user's observations while searching information sources. Ingwersen suggests this may reveal congruence, or lack thereof, with "others" constructs. Sutton (1993) adds that language constraints are most difficult to recognize when the social context is familiar. Although Sutton was referring to the participant in his discussion, it has most application in this instance to item development, in that while writing items experts may inadvertently use language unfamiliar to students.

The six students met individually with the researchers twice over a period of a week. At the initial meeting, students were informed that the purpose of the exercise was to "test the test," rather than testing their knowledge of the topic. Students were handed the test questions and responses and asked to read them aloud and "talk through" their understanding of the question and their reasons for choosing or discounting a response. Little probing was needed as students quite freely and explicitly articulated their thinking processes as they read aloud each item.

Student comments generally addressed item writing construction and specific word interpretation. Overall, several students thought that too much information was included in the item stems or responses, which served to hide key words like "article" or "short paper." Based on findings from their study on poor item-writing practices, Board and Whitney (1972) reported that extraneous material in item stems often served to make the test item easier for poor students, but more difficult for better students. Therefore, item stems were streamlined as much as possible, and item context shortened.

Students also had difficulty interpreting some terms, like "focusing," "blogging," and "controlled vocabulary." To accommodate Clark and Watson's (1995) suggestion that language used in item writing be simple, straightforward, and at an appropriate reading level for the target audience, the terms were changed to clearer, more familiar words. Ultimately, the think-aloud protocol employed in the one-on-one testing was an essential step that served to identify language and conceptual constraints and to clarify items.

Upon completion of the one-on-one testing and receipt of content reviewers' comments, the UCF project team further revised items and formatted them for survey. This portion of the project was completed during August and early September 2004, during semester break.

Small group testing. Small group testing was conducted September 13 and 16, 2004, after the start of the fall semester and between a series of three hurricanes that significantly impacted the area at the time. Students enrolled in two education classes were asked to complete the pilot test, which resulted in 29 usable surveys. The 62 test items initially used for the small group administration are presented in Appendix F. Student responses were entered into a spreadsheet and item analyses were performed.

Item analysis is generally conducted to assess the quality of items and identify problematic items. Items may be considered problematic for a number of reasons, and can include the following: if items are poorly written which causes students to become confused when responding to them, if they have information imbedded in them that may mislead students, if they do not have a clear correct response or if they have a distractor

that could qualify as the correct answer, if they represent a different content area than what is intended, or if they are biased against a subgroup of test takers (Popham & Husek, 1969; Varma, n.d.). Items such as these are unsatisfactory as they cannot properly discriminate between the more and less knowledgeable test-takers. Qualitative checks, such as expert review and think aloud procedures, can help identify problems, but it is possible that some items can slip through as they appear perfectly acceptable on the surface, but are problematic when examined with statistical analysis.

Distractor response analysis was performed, and difficulty levels and discrimination indices calculated for each item. Distractor analysis entailed reviewing the percentage of students selecting wrong choices, and deleting distractors that are too deceptive or not chosen at all. The most problematic items were those with multiple correct responses. Items with multiple correct answers were allowed by Project SAILS guidelines, but analysis revealed that these items were exceedingly difficult and there was question of their ability to discriminate knowledge of the measured concept. The UCF project team recommended that each of these items be revised to include only one correct response. The following example illustrates an item with multiple correct responses and the same item revised.

> Item as written for test:
> You want to locate information on student plagiarism. If you type in the term "plagiarism" as a keyword search, what part of the record is being searched? CHECK ALL THAT APPLY.
> a. abstract or contents field
> b. author field
> c. subject headings field
> d. title field

Suggested revision:
You want to locate information on student plagiarism. If you type in the term "plagiarism" as a keyword search, what part of the record is being searched? CHECK ONLY ONE ANSWER.

a. all parts of the record
b. the abstract and title fields only
c. the author and abstract fields only
d. the author, abstract and subject headings field

Difficulty levels reflect how easy or difficult an item is for the group as a whole, and are computed by dividing the number of correct answers by the number of total answers, then multiplying the value by 100. Non-intuitively, the higher the difficulty level, the easier the item is for the group to answer correctly. Item difficulty levels from the small group administration were well spread. One item had a difficulty level of .07, which meant that two people answered the item correctly, to a high of 1.00, which meant that everyone in the group answered the item correctly. The remaining items had difficulty levels that ranged throughout the scale.

Item discrimination index statistics were computed by looking at the proportion of examinees who selected the correct choice in comparison to their performance on the test as a whole. The point biserial statistic, which is a correlation coefficient, is most often used to calculate the item discrimination index. Varying between 1 and -1, the point biserial statistic indicates the extent to which an item answer correlates to test performance as a whole for the top scoring 27% and the bottom scoring 27%. Top performing students are usually expected to answer difficult items correctly more often than low performing students, a positive point biserial value indicates that this expectation is occurring. As a result, items with positive correlations are sought, and those with correlations of .2 or better are considered to be performing as expected.

Items with negative values signify that lower performing students are choosing the correct response more often than high performing students. Negative coefficient values may pose a problem as it could mean top students are not able to effectively discriminate among distractors. Simon (1969) proposes that negative coefficients can occur when the relationship between performance and knowledge is nonlinear. Popham and Husek (1969) offer that negative statistics may also be attributable to deficiencies in instruction. However, these researchers concede that negative coefficients are likely due to item deficiencies. Review of the item discrimination index data revealed that four of the 62 items discriminated negatively. The negatively discriminating items were carefully scrutinized and suggestions for improving their performance were submitted to Project SAILS.

Summary of Phase I

Despite continuous revision, a question remained regarding the ability of some items to discriminate knowledge among better performing and lower performing students. These items were flagged for Project SAILS review and the UCF project team recommended they be used judiciously or not be included in the test bank. A final report was submitted to Project SAILS at the end of September 2004. This fulfilled the terms of the fellowship and culminated Phase I of the study. Project teams retained rights to further use of the items and publication relating to them.

Phase II of the Study – Instrument Development

As previously mentioned, Phase II of the study extended the first phase, which described procedures to fulfill the Project SAILS fellowship to develop a bank of test items for assessment of information skills specific to the field of education. Phase II of the study explains additional procedures to develop an instrument. These procedures were comprised of selecting and revising items from the bank for the test, designing the test and writing demographic and non-content questions, administering the test and analyzing results, and examining additional calculations.

Item Reduction

From this point forward the researcher worked independently. The first priority was to select items from the 62-item test bank to populate the single test form. Researchers (cf., Clark & Watson, 1995; Davis & Diamond, 1974; Popham, 1974) caution this is another critical point in development of single form tests, especially those that test subdomains within the broad construct. With 62 items from across four content clusters, the item bank offered sufficient representation of the breadth and depth of information literacy skills to be assessed as defined by the NETS*T. Popham (1974) advises that all objectives should have representation on a test, but acknowledges that test developers may be practically constrained from including all items. In this likely event, Clark and Watson (1995), Popham (1974), and Skager (1974) suggest sampling across all significant categories to achieve representation.

Additional guidelines exist for determining optimal test length, which is defined as a test where students' skills levels are correctly identified. Novick and Lewis (1974) offer that the minimum number of items to include on a test depends on how test results are used to make decisions about individual students. For instance, if the test plays a major role in making significant decisions about students, such as summative grades or entry into the next level, test accuracy, and therefore test length, is of particular importance. Lower stakes tests or assessments that rely on multiple measures can require significantly fewer items. To arrive at some idea of the number of items to include on a test, Novick and Lewis looked at how accurately tests reflected students' skills levels with as few as eight test items and as many as 22 items. They found that the accuracy of a test improves as the number of items increases from eight to 22. The current study followed suggestions from the literature, with five to six items being selected from each content cluster for a total of 22 content items.

In addition to content cluster representation, inclusion decisions were based on the range of skills covered by the item, relation of the item to the objective, and the perceived importance of the skill measured by the item. Item analysis and content reviewers' scores, along with the range of difficulty levels and point biserial correlations analyzed from small group results, were also considered in determining which items would be included on the test. However, as content validity is recognized by many researchers as the most fundamental validity for criterion-referenced tests (Horn, 1966, 1968; Popham & Husek, 1969; Simon, 1969), reviewers' scores were accorded more consideration than item analysis statistics.

Item Revision

Project SAILS guidelines provided a lot of latitude in determining what was considered "acceptable" item writing. For instance, items could have varying numbers of responses or multiple correct answers. For the final version of the test, items were revised based on criteria for item writing that were uncovered in the literature, expert reviewers' comments, and students' observations. Test items were rewritten to reflect a single correct response after content reviewers expressed concern regarding items with multiple correct answers. Further, a consistent number of responses were used throughout the instrument. In a study investigating the effect of item-writing flaws on test reliability and validity, researchers found that poor item-writing practices can obscure or attenuate differences between good and poor students (Board & Whitney, 1972). Item-writing criteria offered by Board and Whitney include using simple, straightforward language, avoiding "all of the above" and "none of the above" responses, making responses similar in length, clarifying text of stems and responses, and randomizing response order by alphabetizing the first word. Study procedures followed these recommendations.

Test Design

In practical application, test administration would most likely occur during class time. Consequently, test length was a chief issue and a balance was sought between time to take and score the test and test length. Multiple-choice items can be answered and scored more quickly than short answer, essay responses, or observation of actual

behavior, so the multiple-choice format endorsed by Project SAILS was retained for actual instrument development. A perceived limitation of multiple-choice tests is their ability to discriminate among skills levels. This is attributed to correct responses due to guessing that are more difficult to discern among multiple-choice items when compared with constructed response items. However, multiple-choice tests are the preferred format for many library and information literacy assessment tests. For example, researchers at James Madison University, which has been conducting large scale assessment of students' library skills levels since 1989, report that multiple-choice tests are a good way to evaluate program effectiveness (Cameron, 2004). This view has been supported by Project SAILS administrators and other studies, which have likewise found that well constructed multiple-choice tests can acceptably discriminate among comprehension levels (Anderson, 1974; Hakstian & Kansup, 1975; Wilbur, 1970).

Demographic items, addressing such topics as gender, ethnicity, student classification, and length of enrollment, were also developed. Other non-content related items included questions pertaining to students' perceptions and experiences with libraries and information use. Differences in scores may be affected by any number of factors represented by the demographic or non-content area items. For example, items inquiring about exposure to and type of library-provided instruction may provide some insight on the number and dispersion of students who receive library instruction as well as effectiveness of different aspects of the instructional program. Data from these questions were used as the basis of independent variables analysis described in the Results chapter.

Population and Setting

Undergraduate education students were chosen as the population for this study based on the call by ACRL and NCATE. There is particular need for colleges of education to produce information literate students who can access the literature to inform their professional practice and who can, in turn, model and teach these skills to their students. Although the test could also be used with graduate students in education, there is expectation that this population of students should be exposed to a wider variety of information databases and sources. Therefore, the test developed for this study may not assess the full range of skills and knowledge expected of graduate students. A valid, reliable, and current instrument that measures information literacy skills levels of undergraduate education students would be useful to education faculty and librarians working with that population.

Students who participated in testing were from a public, metropolitan university with enrollments of over 40,000 students. Final fall 2004 headcount data reported 3,053 undergraduate students enrolled in the College of Education (UCF Office of Institutional Research, 2005). Of the 3,053 students, 83.85% were female and 16.15% male. Regarding ethnicity, 80.56% were identified as White, 9.62% as Hispanic, 7.39% as Black, and 1.96% as Asian or Pacific Islander. Teaching degrees are offered in early childhood and elementary education, art education, English and foreign languages, mathematics education, physical education, science and social science, and vocational education. Students have access to a number of electronic general academic and

education databases and are expected to attend planned library instruction sessions and orientations as they progress through their program.

Sampling Technique

The researcher met with representatives from the university's Operational Excellence and Assessment Support (OEAS) office in September 2004 (after several canceled appointments due to inclement weather) to discuss test administration support from the office. OEAS agreed to host the electronic version of the informed consent form and instrument and collect responses to the test. Initially, study plans had included use of an Office of Institutional Research-supplied sampling frame to obtain representative demographic and enrollment samples, but OEAS protocol precluded use of the sampling frame. Instead, the researcher and OEAS representatives agreed to extend an email invitation to complete an electronic version of the test to the identified population of undergraduate students majoring in education.

For surveys and tests that measure in the affective domain, a general rule for establishing sample size is five times the number of variables on the test (Hatcher, 1994). No similar recommendations were found for criterion-referenced, cognitive assessment sample sizes, so estimates for survey research and affective tests were relied upon. With 22 content items, a minimum sample size of 110 respondents was sought.

Test Administration

Several factors were taken into consideration when converting the test to electronic format. To ensure maximum response rate, Dillman Tortora, Conradt, and Bowker (1998) suggest using a Web-based survey with a front-end that is less graphics-heavy which generally results in higher response rates than designs with more programming and graphics. Consequently, the Web-based version, with black text on a white background and a minimum of buttons and graphics, was designed to closely approximate the print test. Dillman (1999) adds that the length of an instrument has an inverse relationship to response rate; participant completion of the test lowers as test length increases. This concern tied into the balance sought between an adequate number of test items and test length, with the researcher concluding that deleting any content or demographic items would seriously compromise the validity of the test.

Electronic. Participants were asked to respond to a 35 item, multiple-choice format test that contained 22 content questions and 13 demographic and self-percept questions. The electronic version of the test contained the same items as the print version, but administration procedures differed. Students who completed the electronic version did so upon email invitation. The email described the library's effort to investigate the effectiveness of its instruction and asked for students' assistance with the evaluation. The Dillman (1999) five-contact method was used, and all students received four follow-up emails after the initial invitation. For the electronic version, students clicked on a link from the email request that took them to the informed consent form. Clicking on a button

embedded in the form signified consent and led students to the actual test. The test was completed and submitted online. The testing window remained open from late October through mid-November, a period of three weeks.

Despite actions followed to enhance response rate, the number of responses was exceptionally low. Although up to 40% of contacts can be expected to respond to a request to complete an electronic survey (Dillman et al, 1998), only 3.0%, or 92, of the population contacted submitted surveys that were usable. This exceptionally low response rate may have been influenced by a number of testing and environmental conditions. As previously mentioned, the electronic version of the test was open for three weeks. Although severe weather affecting the area was most pronounced mid-August through mid-September, subsequent recovery efforts and attempts to get back to some sense of normalcy most likely relegated completion of the survey to very minor importance.

Administration of the test also was delayed due to power outages and school closings. This placed the testing period at the time most likely for midterm exams and holiday plans. Further, the length of the test may have been a deterrent to its completion. A number of students started answering the items, but submitted the test before completion. Of the 127 tests returned, only 92 were sufficiently complete to use. Finally, several mistakes were made in administering the test. The first mailing of the invitation included a non-working link to the test. Subsequent mailings were either erroneously sent to all graduate and undergraduate students or referred students to the same non-working link. After the test site closed, response data were sent electronically to the researcher.

Print. The number of usable responses received made it apparent that a second administration of the test was needed. Failure to achieve the minimum target sample of 110 usable tests provided an excellent opportunity to rethink the impact of administration format on student performance. It is highly likely both print and electronic versions of the instrument would be used in practical application, and the utility of investigating whether the various test versions impacted student performance was deemed important. Therefore, the number of additional tests needed to ensure an adequate sample size for analytic procedures rose from the initial deficit of 18 to a similar number of results as received from the electronic version of the test. A second administration of the test, this time in print, was planned for February 2005.

Undergraduate education students were notified of the opportunity to take the test via a sign placed in a busy lobby of one of the education buildings. Students who responded to the call were orally informed of the research project and the approximate time it would take to complete the test (30 minutes). Test packets were prepared in advance, and consisted of a copy of the test, a Scantron answer sheet, a cover sheet explaining the purpose of the research, and two copies of the informed consent form. One copy of the consent form was to document consent for the researcher, the second copy was offered to students for their records. Students completed the test in a branch of the library.

Students' university-assigned personal identification numbers were collected on the informed consent form and the Scantron, but most were stripped after responses were entered into the database. Only identification numbers of students who indicated they

were interested in a brief follow-up test were retained until completion of the second test. Several participants expressed great interest in the test and noted the need for such an instrument. An additional 80 tests were completed, bringing the total to 172. See Appendix G for the test and scoring key.

Further Development

The issue of establishing validity of criterion-referenced tests that rely on score variability procedures was addressed in Chapter Two; however, a brief discussion of a related debate is warranted. The concept of validity with respect to test development has many nuances, yet an alarming number of researchers apparently believe that construct validity can be established simply by reporting results of factor analysis procedures (Clark & Watson, 1995). Indeed, in an historical overview of factor analytic evidence, Thompson and Daniel (1996) point out that factor analysis and construct validity have long been associated with each other, and have even been erroneously conceived and reported as the same concept.

Other researchers envision a much more interrelated and dependent model, one that seeks a number of procedures to build evidence of validity and recognizes construct validity as the overarching validity. Construct validity is often characterized as the degree to which the construct has been successfully operationalized, or the demonstration that a test measures the construct it claims to be measuring. Remaining validity procedures are generally considered to be supporting facets of construct validity (cf., American Psychological Association [APA], 1999; Brown, 2000; Clark & Watson,

1995). Therefore, criterion-related validity, content validity, and even some reliability measures serve to verify a more unified and central view of construct validity. Clark and Watson add that construct validity "lies at the heart of clinical utility of assessment" (p. 310) and its establishment should be demonstrated from a number of perspectives and an accumulation of evidence.

Criterion-related validity. Criterion-related validity, or the degree to which the measure is deemed accurate by comparing it to another measure or procedure that has been demonstrated to be valid, was established by comparing test answers to actual performance on related library and information-seeking tasks. The in-library test was developed from the written test. The eight items for the in-library test were selected from the written test based on the criteria of ease of performance in the library and representation of the four content clusters. Two items were selected from each of the content clusters.

Selected items were checked for item difficulty levels and point biserial statistics and found to represent a broad range of the two measures. Item difficulty of the subtest items ranged from .32 to .68, compared to .32 to .89 on the full test, so selected items were slightly more difficult when compared to the full test. Point biserial statistics for the subtest ranged from .08 to .36, compared to .08 to .44 for the full test, indicating item discrimination was representative of the whole set of items.

Ten participants were selected from the pool of test-takers who had indicated they were interested in participating in a follow-up test. Five of the students had test scores at or below the mean score of 11.9, or 54%, while the remaining five had test scores above

the mean. Test scores of participants ranged from a low of 8, or 36%, to a high of 19, or 86%. Participants were scheduled for one-half hour time slots to complete the in-library test.

To facilitate consistent administration of the test, scenarios for each item were devised and language asking the question was decided in advance. Prior to each student taking the in-library test, all books were placed in the same order and computer screens set to the same site. Questions were grouped for maximum efficiency. For example, all print source-related items were asked together and all questions that entailed computer use were grouped together.

The following are examples of a written test item and the corresponding in-library test item. For the in-library test, the scenario is described in the brackets and verbal instructions are indicated by quotation marks. See Appendix H for the complete in-library test.

Written Test:

Item 18. You have a class assignment to investigate how group work impacts student learning. A keyword search in ERIC on "group work" has returned over 600 items. To narrow your search, which of the following steps would you next perform?

a. add "impacts" as a keyword
b. add "student learning" as a keyword
c. limit search results by date
d. limit search results by publication type

In-library Test:

Item 18. [Show student *ERIC* database search screen, with search strategy showing the results from a keyword search on *group work*.]
"You are searching how group work impacts student learning. What is the next logical step when you get this many results?"

Written Test:

Item 20. Your professor suggested you read a particular article and gave you the following citation:
Shayer, M. (2003). Not just Piaget, not just Vygotsky. *Learning and Instruction, 13*(5), 465-485.
Which of the following would you type into the library's catalog to locate the actual article?
a. author search: Shayer
b. journal title search: *Learning and Instruction*
c. journal title search: Not just Piaget, not just Vygotsky
d. subject search: Piaget and Vygotsky

In-library Test:

Item 20. [Show student library catalog search screen. Hand student following citation: Shayer, M. (2003). Not just Piaget, not just Vygotsky. *Learning and Instruction, 13*(5), 465-485.]
"Type in what you need to locate the item."

The in-library test was conducted in a branch of the main library and administered by the researcher. This phase of the testing occurred anywhere from 14 to 20 days after students completed the written test. Students were tested individually. After students arrived at the testing site, they were greeted and thanked for their participation. Students were given instructions to complete the task requested. When answers were ambiguous, or students indicated they knew what to do but could not execute the procedure, the researcher probed with open-ended questions. Time spent with the students ranged from 15 minutes to 30 minutes. Although the researcher did not offer to discuss results of the in-library test, students were asked if they had any questions after completing the test. Six of the ten students asked specific questions about individual items or overall results, or mentioned the need for more instruction in this area.

Following a scoring protocol established by Morner (1993), a coding sheet was developed that listed the item answer on the initial test and the in-library item answer. Responses to the items were compared between the written and in-library test, with items noted that changed from one test administration to another.

Factor analysis. Four content clusters were identified from the NETS*T standards, and were described as: identifying, evaluating, and selecting finding tools, demonstrating knowledge of general search strategies, evaluating and selecting sources, and demonstrating knowledge of legal and ethical practices. Clark and Watson (1995) suggest that factor analytic procedures are most appropriately used when the target construct is conceptualized as being multidimensional. Therefore, factor analysis was performed to further explore construct validity by investigating the extent to which the content clusters operationally represented unique factors. Generalized least squares factoring was the extraction method, and as factors were believed to be unrelated, orthogonal rotation was deemed appropriate.

Reliability

Reliability refers to the extent to which the test is likely to produce consistent scores. These procedures infer the consistency with which the test is measuring the intended construct. Instrument stability and internal consistency are two procedures often used as evidence of reliability in test development. Instrument stability procedures are conducted to indicate consistency of scores over test administrations. Generally, this

consistency is determined by readministering the test to the same population or a subset of it over a lapsed period of time.

The internal consistency value reveals how items cohere or relate to each other. This statistic reflects several characteristics. Foremost among them is the assumption that the greater the number of correlated items and the stronger the relationship, the greater the reliability. Other generally accepted characteristics are that longer tests increase reliability, and more diverse subject matter lowers the reliability. Therefore, a high internal consistency value is achieved when students who answer items correctly are more likely to answer other items that cover the same content correctly. Conversely, a low score may be interpreted as meaning that items tend to be unrelated to each other in terms of who answered them correctly. As with construct validity and factor analysis, the reader often finds studies reported in the literature whereby the concept of internal consistency is reported as instrument reliability.

Horn (1968) explains that reliability and internal consistency are not identical. Horn further differentiates between the two concepts by explaining that reliability is the ratio of true score variance to observed score variance and internal consistency is how well items cohere together. Additionally, Horn (1966) maintains that internal consistency may be a secondary consideration for criterion-referenced tests and can actually be a counter indication of a test's adequacy. He does not question that an instrument should minimize error of measurement or that internal consistency can be indicative of reliability

in the broad sense.[5] Instead, Horn (1966, 1968) argues that a criterion-referenced test does not need to have many of the properties deemed desirable for predictive measures. To illustrate, Horn proposes that if there are a distinct number of areas of a particular construct, and each are represented by one question, one could expect only random correlation among the items. It is widely acknowledged that well constructed tests covering topics with a discrete body of knowledge, such as vocabulary or mathematics, are more homogeneous (higher reliability) than well constructed tests of more heterogeneous topics (see Kehoe, 1995). Kehoe (1995) adds that homogeneity varies from discipline to discipline. Therefore, disciplines with less homogeneous course content should perhaps be satisfied with lower test homogeneity.

By applying this example to information literacy assessment, it becomes readily apparent that a student may have knowledge of some aspects of the construct, but not all of them. Information literacy is a very heterogeneous and non-hierarchal construct (Radcliff, 2005). It is reasonable to expect that students who have deeper knowledge of some aspects of information literacy may also be schooled in other related areas, but consistently correct answers could also indicate the universe of information delineating the construct is not being tested. It is this type of case that leads Horn (1968) to reason that high internal consistency could be a counter indication of validity. For these reasons, Horn and Popham and Husek (1969) conclude that internal consistency does not

---

[5] Ebel (1968) agrees that in theory reliability and internal consistency are two distinct concepts, but reminds us that practically, reliability is estimated by the same correlation coefficient that measures internal consistency.

constitute primary evidence in evaluating a criterion-referenced instrument. Instead, they suggest the primary concern is ensuring that experts agree the content of the scale is appropriate.

Stability. Instrument stability was measured by a test-retest procedure which entailed administering the written test twice after an approximate two week interval. During initial administration, students checked a box on the volunteer consent form indicating their willingness to participate in further testing. Fifteen students were contacted and asked to take the test a second time. All agreed. However, four were not available during the testing window. Administration of the second test was conducted under conditions as closely approximating the first test as possible. Students were given the same test form and instructions, and testing took place in the same facility. Approximately two weeks lapsed between test administrations for the group.

Student personal identification numbers were used to match tests from the two administrations, after which the identification numbers were stripped and tests were renumbered (e.g., test 1.a. and 1.b., 2.a. and 2.b., etc.). The number represents the test taker and the letter indicates first or second administration. Scores from the initial test and the retest were tabulated and results of the 22 content questions from 11 participants were compared.

Internal consistency. Cohen and Cohen (1983) note that the purpose of reliability procedures is to determine the correlation between the observed score and the true score. The debate as to the importance of, or even the desire for, high internal consistency

coefficients for criterion-referenced tests has been discussed. However, these values can also indicate measurement error, so results were submitted to internal consistency coefficient calculations. The Kuder-Richardson 20 (K-R 20) formula is used to compare dichotomously scored data with continuous data, and is analogous to the alpha statistic, which is used to calculate internal consistency values of scales when compared variables are continuous. As test items were either correctly or incorrectly answered, responses were dichotomously scored and were submitted to the K-R 20 calculation for item-subscale correlations.

Passing Score Calculation

The rationale for establishing cut scores was discussed in Chapter Two; procedures are presented here. Briefly, Berk (1986) recommends decisions regarding competency be based primarily on judgment methods and supplemented by empirical methods. Pulling components from the Angoff-iterative and informed judgment methods, Berk suggests a two step process. A panel of judges initially reviews and rates the items, then are allowed to adjust their decisions based on item statistics. Several factors were considered when deciding which method was most appropriate to calculate the cut score for the B-TILED. In addition to Berk's recommendation, another factor for consideration was the anticipated skills level of the test takers. The test group was very heterogeneous in both skills levels and exposure to library instruction, and it was believed the group would have a lower score, on average, than an instructed group. Therefore, a

method that relies primarily on statistical procedures was discounted in favor of a judgmental approach.

For this study, five judges were asked to estimate the percentage of minimally knowledgeable candidates who would know the answer to an item. After judges made their first best attempt, they were provided item difficulty levels and demographic information from the field test. Judges then had the opportunity to revise their estimates. To this point, the procedure could be classified as a two-step Angoff-iterative method. However, one additional step was added.

A cut score is an observed score, and therefore consists of the true score and measurement error. In his review article, Berk (1986) reports that the cut score setting method should identify measurement error then accommodate the difference between the observed score and the true score. Fortunately, measurement error can be estimated. Although the two-step Angoff-iterative method did not include this procedure, the adjusted/modified multiple-choice Angoff method makes an allowance for unreliability. This procedure adjusts the item percentage by the standard error of the mean in the direction necessary to minimize either false competence or false incompetence errors. This procedure was deemed beneficial to the cut score calculation and was added as a final step. As information literacy skills are considered relatively low stakes at this point, the standard error was deducted from the cut score calculation.

Summary of Phase II

Procedures followed in development of an initial item bank were described in Phase I of the study. Development and design of the test and its administration were included in Phase II. Upon completion of the electronically-administered test, data were forwarded to the researcher in an Excel file. Results gathered from administration of the written test were appended to the original Excel file. Data were exported to SPSS version 10.0 and various statistical calculations were performed. Test items were submitted to analysis. Procedures included analyzing distractors for plausibility and calculating item difficulty levels and discrimination indices.

A subtest of the B-TILED, with ten participants replicating tasks associated with test items in an authentic environment, was administered to check criterion-related validity. Factor analysis of the scale and content clusters was performed. Reliability procedures consisted of a test-retest to investigate score consistency and stability, and internal consistency calculation. To measure stability, eleven students were administered the same test form twice and results were analyzed. Internal consistency was calculated using the K-R 20 formula for item-subscale correlations. The test was also submitted to variations of the Angoff method for establishing a passing score. Results of these procedures and statistical measures are described in Chapter Four. A summary timeline of the study is presented in Appendix I.

CHAPTER FOUR
RESULTS AND ANALYSIS

This chapter describes the results of various procedures and calculations performed as part of the study. The chapter begins with a recap of procedures followed for test development, then provides descriptive statistics of the sample and the test. Frequencies of demographic variables, including gender, ethnicity, student classification, and length of enrollment at the institution, are used to describe the sample. Item level data were analyzed and distractor plausibility, item difficulty levels, and discrimination indices are reported. Statistics that describe the test include distribution, averages, and range of scores. Test data were also analyzed by respondents' characteristics and various validity and reliability procedures, and calculations are reported. The chapter concludes with the results of the passing score calculation.

Procedures for Test Development

Upon completion of the Project SAILS fellowship to create items for a test bank, development of an instrument that could be easily administered and scored by faculty began. Criteria used to reduce the 62 items developed for Project SAILS to the 22 items included on the B-TILED were based on alignment of the NETS*T standards with existing ACRL objectives, rating scores from five content experts, and results of a one-on-one think aloud protocol and subsequent small group pilot testing. Thirteen demographic and self-percept questions were added to the 22 content questions; the final

version of the test totaled 35 items. The test was administered to 172 education students enrolled a large urban university and results were submitted to analysis.

Descriptive Statistics of the Sample

Undergraduate students enrolled in a teacher education program at the university were contacted via email and asked to complete an electronic version of the instrument. This administration netted 92 usable surveys. The extremely low response rate resulted in the need to supplement the sample, and an additional 80 print-administered surveys were collected. For the electronically-administered test, the population of undergraduate education students was contacted five times over a five week period. Correspondingly, all students entering the lobby of the education building during any of the administration times of the print-based test were also invited to participate.

Although all education students were contacted and asked to complete the test, not all chose to comply. This led to the possibility of non-response error, which is the extent respondents differ from non-respondents. One method of addressing non-response error is to compare characteristics of respondents to characteristics of the population (Groves, 1989). Final fall 2004 headcount data indicated 3,053 undergraduate students were enrolled in the College of Education (UCF Office of Institutional Research, 2005). Of the 3,053 students, 83.85% were female and 16.15% male, and 80.56% were White, 9.62% Hispanic, 7.39% Black, and 1.96% Asian or Pacific-Islander. Combined print and electronic administration of the test resulted in 172 usable surveys. Of the participants who completed the survey, 136 were female (80.00%) and 34 were male (20.00%); two

students did not respond to the item. 167 participants responded to the question of ethnicity, with the majority White or European-American (81.39%), followed by Black or African-American (8.14%), Hispanic or Latino (7.56%), and Asian or Asian-American (2.32%). One person who listed "other" reported ethnicity as Arab.

Although the sample was not randomly generated, gender and ethnicity proportions of the sample were similar to the population of undergraduate teacher education majors at the institution. Statistics that describe the sample were further analyzed and reported in the section Analysis of Test Data by Respondents' Characteristics located toward the end of the chapter. Table 1 presents summary statistics for the population and the sample.

Table 1: Characteristics of Population and the Sample, by Percent

	Category	N	Population Characteristics (Percent)	Sample Characteristics (Valid Percent)
Gender	Female	136	83.85	80.00
	Male	34	16.15	20.00
Ethnicity	White or European-American	137	80.56	82.03
	Hispanic or Latino	13	9.62	7.84
	Black or African-American	12	7.39	7.19
	Asian or Asian-American	4	1.96	2.40
	"Other"	1	0	.59

Descriptive Statistics of the Test

The mean raw score of the sample was 11.97 (SD=3.74). The standard error of measurement rate was .28, which indicates there is a 95% probability that the scores are accurate to .56 points, plus or minus. The frequency distribution shown in Table 2 reveals scores ranged from 2 to 20, out of a possible 22.

Table 2: Test Score Frequency Distribution

Score	Frequency	Percent	Cumulative Percent	Percent of Maximum Score
2	1	.6	.6	9
3	1	.6	1.2	14
4	1	.6	1.7	18
5	5	2.9	4.7	23
6	5	2.9	7.6	27
7	6	3.5	11.0	32
8	16	9.3	20.3	36
9	11	6.4	26.7	41
10	9	5.2	32.0	45
11	24	14.0	45.9	50
12	17	9.9	55.8	55
13	17	9.9	65.7	59
14	12	7.0	72.7	64
15	15	8.7	81.4	68
16	12	7.0	88.4	73
17	6	3.5	91.9	77
18	7	4.1	95.9	82
19	6	3.5	99.4	86
20	1	.6	100.0	91

The distribution of scores is fairly normal, with 46% of respondents falling into the midrange of 10-14, which closely approximates the second and third quartile. Figure 1 is a graphical representation of the distribution of scores. Additional descriptive statistics for the test are presented in Table 3.

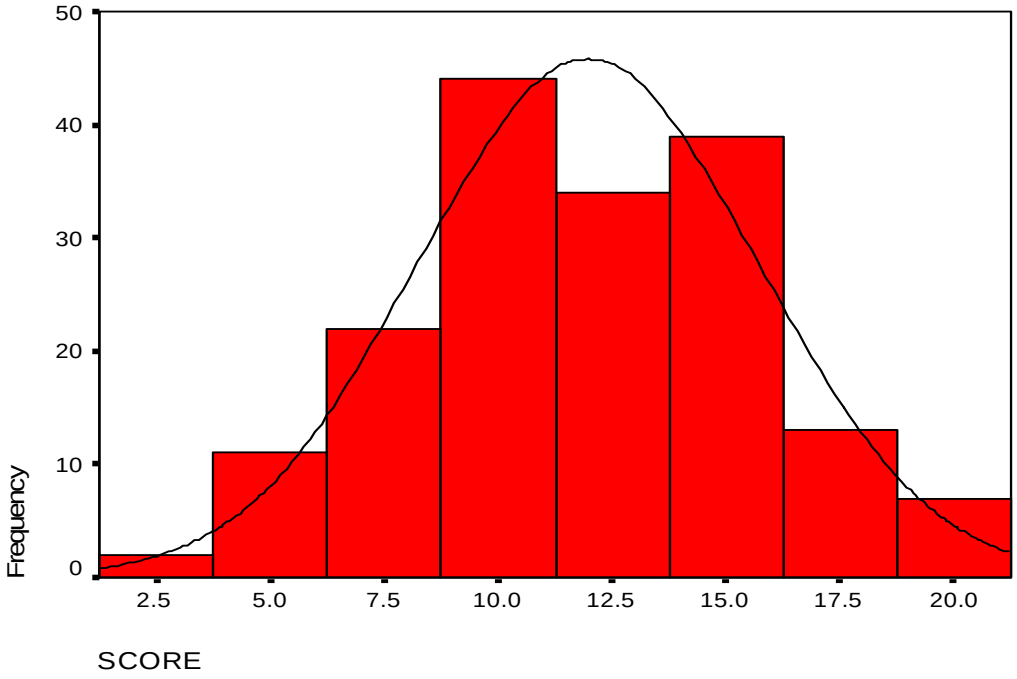

Figure 1: Frequency Distribution of Test Scores

A question still exists whether student test performance is affected by administration mode of the instrument (Russell & Haney, 1997). Fairly equal group sizes allow for comparison of data from print and electronic administration. Table 3 displays descriptive statistics for student test scores by administration mode and for the total sample. Despite a one point difference in mean scores between students who completed the print-administered test and the electronically-administered test, scores were not

statistically significant different at the .05 level. The two subgroups did not differ greatly when comparing the range of scores, the standard deviation, or the standard error of measurement.

Students were not randomly assigned to group, so conclusions must proceed with caution. It may be permissible to infer that modes of administration made little difference for this administration of the B-TILED, but not all test administrations or tests can reach the same conclusion. The K-R 20 value was also relatively stable across test administration modes. However, due to nonrandom selection and the practical difference in scores, future administrators should continue to monitor differences between print and electronic administrations of the test.

Table 3: Descriptive Statistics by Administration Mode and for the Total Sample

	Print Administered	Electronically Administered	Total Sample
Mean Raw Score	11.44*	12.43*	11.97
Mean Percent	51.99	56.53	54.42
Median Pct	50.00	54.55	54.55
Mode Pct	50.59	50.00	50.00
Range Pct	9 to 86	14 to 91	9 to 91
K-R 20	.673	.678	.675
SD	17.08	16.70	16.98
SEM	1.91	1.74	1.29
Number	80	92	172

Note: *p>.05, maximum score=22

Table 4 summarizes item level data, including difficulty and discrimination indices and the response percentage for each test stem. In the table, "item number" reflects the actual test item numbering. Items 1 through 6 and 29 through 35 were demographic or other non-content items. "Correct Answer" refers to the correct item response and "Difficulty" denotes the percentage of students answering the item correctly. "Discrimination" is the item discrimination index, or point biserial correlation, which gives the ratio of high-scoring students who answer the item correctly compared to low-scoring students. "Percent choosing" indicates the percentage of students who chose each response, including the correct answer and distractors. Percentages may not add up to 100 due to rounding.

Difficulty levels ranged widely for the 22 items, from 32% answering item 8 correctly to 89% choosing the correct answer for item 25. This indicates the test contained items of various difficulty levels and that students exhibited a broad range of skills levels. Although it is quite possible for test takers to score in the upper ranges on criterion-referenced tests, Kehoe (1995) suggests that items answered correctly by 30% to 80% of test takers are good target difficulty ranges for discriminating knowledge.

A range of difficulty levels is also found among the four content clusters. The first cluster, identifying, evaluating, and selecting finding tools, contains items with a difficulty range of .32 to .68. The second cluster, demonstrating knowledge of general search strategies, ranges from .39 to .73, while cluster three, evaluating and selecting sources, ranges from .36 to .69 and cluster four, demonstrating knowledge of legal and ethical practices, contains items ranging from .34 to .89.

Table 4: Item Level Statistics

Item Number	Correct Answer	Difficulty	Discrimination	Percent Choosing A	B	C	D
7	C	.49	0.31	21	25	49	5
8	D	.32	0.23	34	10	24	32
9	D	.57	0.24	9	8	27	57
10	A	.41	0.27	41	22	35	2
11	D	.39	0.25	37	19	6	39
12	D	.68	0.36	5	7	19	68
13	B	.65	0.16	30	65	4	1
14	A	.60	0.23	60	14	20	6
15	C	.42	0.08	12	21	42	24
16	B	.59	0.44	14	59	10	17
17	C	.73	0.41	10	10	73	6
18	B	.65	0.21	15	65	12	8
19	B	.36	0.35	9	36	49	5
20	B	.43	0.13	23	43	28	6
21	C	.69	0.36	6	3	69	21
22	C	.57	0.19	6	3	57	32
23	C	.42	0.11	5	35	42	18
24	D	.57	0.27	22	10	10	57
25	C	.89	0.22	5	2	89	5
26	A	.34	0.07	34	9	10	46
27	A	.81	0.19	81	5	8	6
28	B	.42	0.15	18	42	29	10

The discrimination index, or point biserial correlation, compares performance on a given item from top scoring students with performance from students in the bottom group. If all students in the top scoring group choose a correct answer and all students in the low scoring group choose a distractor, then the discrimination index would be 1.0. Negative discrimination values indicate top scoring students are choosing an incorrect

answer, while low scoring students are answering the question correctly (Varma, n.d.). Negatively scored items are generally considered to discriminate among knowledge levels inadequately, so revision is recommended. No negative item discrimination values were uncovered, which indicates that test items discriminated between high and low scores in the desired direction.

Two items, number 15 and number 26, had discrimination values below .10, and two additional items (numbers 20 and 23) had values below .15. Some researchers (Kehoe, 1995; Varma, n.d.) suggest items with a discrimination index below .15 be reviewed and either revised or withdrawn. Popham (1974), however, contends that items may be left in if they are well written. A range of opinions exists among researchers as to whether it is advisable to discard items that appear to be poorly performing. In general, researchers who analyze measurement results from the affective domain suggest deleting low performing items. Those developing cognitive, criterion-referenced tests argue that the researcher consider importance of the knowledge of the objective over performance of the item. After meticulous review, the researcher decided not to delete or revise the items as it was believed the items did discriminate among knowledge levels.

The "percent choosing" columns provided the basis for distractor analysis. Every alternative was chosen at least once, and five items demonstrated a good dispersal among choices with at least 10% choosing each alternative. Distractor analysis was also performed during test development, and served to identify implausible distractors. Continued analysis can inform future revisions of the test. For example, in item 13 response D was chosen only once. A more plausible alternative should be considered for the item.

In summary, the test was administered electronically and in print to a sample of 172 education students. Scores were distributed fairly normally, and ranged broadly, from 2 to 20, out of a possible 22. The mean score for the sample was 11.97, or 54.42%. The K-R 20 calculation revealed a value of .675 and a standard error of measurement of 1.29. Difficulty levels of test items ranged widely and no test items had a negative discrimination value. All test item responses were chosen at least once.

Construct Validity

All introductory statistics textbooks offer a section on validity, and portray it as fundamental to any study. Validity is generally defined as determining whether a test measures what it purports to measure, and most texts go on to suggest that multiple procedures are required to attest to an instrument's validity. Nonetheless, it is not uncommon to see "validity" reported both in the scholarly literature and on numerous statistics Internet sites as a single alpha coefficient. This has led Clark and Watson (1995) to caution that many researchers have a naïve understanding of construct validity. They state, "Construct validity cannot be inferred from a single set of observations…" (p. 310), but instead suggest a number of procedures should be used to provide evidence. Content validity, criterion-related validity, and factor analysis are procedures universally accepted by the research community. Results of these procedures follow.

Content Validity

Content validity is generally defined as the degree to which a test reflects all aspects of the dimension or construct being measured. Linacre (2004) adds that content validity should be used as an initial screening device, and that the procedure should verify that although extraneous material has been omitted, all relevant material is represented. For the current study, characteristics of the construct of information literacy were represented by the ACRL and ISTE standards. These criteria describe what content should be included in information literacy instruction, and the cognitive knowledge students should possess to be considered information literate. Content validity of objective measures is often determined by subject experts who evaluate individual test items and determine whether the items represent the intended objective and construct.

As described in the Methods chapter, five content experts were asked to evaluate each of the items on the criteria of accuracy, clarity, and institutional objectivity. Items were scored on a scale of 0 (low) to 3 (high). Averages of reviewer scores for the 22 test items that were included on the B-TILED are presented in Table 5. As this exercise was conducted during the initial phase of the test, reviewers were sent all 62 test items. The item numbers in the following table identify the item as it appears on the final scale. Appendix J summarizes individual reviewer ratings for all 62 items.

Table 5: Mean Average of Reviewers' Rating, by Item

Item Number (B-TILED)	Accuracy	Clarity	Institutional Objectivity
7	2.6	2.6	3.0
8	2.6	2.8	3.0
9	1.8	1.4	2.8
10	2.4	2.4	2.6
11	3.0	2.0	2.2
12	2.8	2.4	2.8
13	2.8	2.4	2.2
14	2.8	3.0	3.0
15	3.0	3.0	3.0
16	3.0	2.6	3.0
17	3.0	2.8	2.6
18	2.8	2.8	3.0
19	3.0	3.0	3.0
20	3.0	3.0	2.6
21	3.0	3.0	3.0
22	2.2	2.0	3.0
23	2.8	2.2	3.0
24	2.4	1.8	3.0
25	2.8	2.6	3.0
26	2.2	2.6	3.0
27	2.6	1.6	3.0
28	2.2	2.4	3.0
Mean	2.67	2.47	2.85

For the criterion of accuracy, reviewers assigned fairly consistent ratings across the items. When reviewers were asked to evaluate each item on a scale of 0 (low) to 3 (high) in regards to how accurately the item described the objective, all five reviewers scored items at a 2 or 3 level 95% of the time. The average score by item of all 5 content experts ranged from 1.8 to 3.0, with a mean score of 2.67. Item clarity of the 22 items retained for inclusion on the final test was also fairly high. Of the 22 items, 19, or 86%,

received an average score of 2 or more. Three items that received a rating lower than 2 were reviewed and revised. The mean score for the 22 items was 2.47. As the test was devised to be used across multiple settings, institutional objectivity of the item was another important quality. Using the same 0 to 3 scale for accuracy and clarity, the experts scored institutional objectivity very highly. All item average scores were 2.2 or higher. The mean average for objectivity across all items was 2.85.

Content validity, as determined by a panel of five experts who have worked extensively with education students in the context of their information-seeking, was scored consistently excellent. The accuracy of individual items as they relate to an identified information literacy learning objective, the clarity of the items as written, and their institutional objectivity were all corroborated by the content experts.

Criterion-Related Validity

Criterion-related validity procedures are used to determine how well the test compares to another measure or predicts ability of the construct being assessed. This check is frequently performed by comparing participant performance on one measure with their performance on another. For this study, criterion-related validity is concerned with measuring students' abilities to execute information literacy skills in an authentic environment. Procedures included administering a test comprised of a subset of items from the B-TILED. To distinguish between the two tests, the original, full length test will be referred to as the *written test* and the subtest administered in the library will be referred to as the *in-library test*.

The in-library test was developed and administered based on protocols established by Morner (1993) in her development of the Morner Library Research Skills Test. Results from the written test were compared to results of the in-library test to establish the degree of criterion-related validity of the written test. Ten student participants replicated the written test with an in-library test using a subset of the items. Five students had test scores below the mean score of 54% on the written test and five had scores above, with test scores ranging from 36% to 86% correct. Each student answered eight items from the library test that corresponded to eight items from the written test.

Results from both tests were compared. Table 6 reports item comparison results for three categories: the number of items with no change, the number of items correct on the written test but incorrect on the in-library test, and the number of incorrect written test items compared to correct in-library test items. When comparing results among the eight items on each test, 78.8% of the answers did not change, 12.5% changed from correct to incorrect, and 8.7% changed from incorrect to correct. These results suggest a fairly high correspondence between the tests, which is an indication that the test reflects students' real performance.

Further investigation of higher and lower scoring students was conducted, and the comparison revealed a slight difference in performance between the two groups. The five students who scored below the mean had ten answers that changed on the two test administrations while students who scored above the mean had seven answers change. The students scoring above the mean had slightly more stable scores, perhaps suggesting that increased variability of the lower scoring students was due to guessing answers.

Table 6: Comparison of Scores between Written Test and In-library Test

Student	Number of Items with No Change	Correct Written Test to Incorrect In-library Test	Incorrect Written Test to Correct In-library Test
1	7	1	0
2	6	2	0
3	7	1	0
4	6	0	2
5	5	2	1
6	8	0	0
7	7	0	1
8	6	2	0
9	7	1	0
10	4	1	3
Total	63 (78.8%)	10 (12.5%)	7 (8.7%)

An analysis of items that differed between the written test and the in-library test revealed that, in most cases, only one or two students changed answers for each item. In only two cases students changed their answers three or more times. Item 12 accounted for three changes in answers, and item 13 had five changes. For item 12, students were asked to select the source which was first to publish research studies in education. Eight out of ten students who took the in-library test chose response D, professional conferences and education journals, on the written test. However, three students who chose the correct answer on the written test opted for the alternative of education newsletter during the in-library test. The researcher chose a book, an education journal,

and an education newsletter that all contained the word "research" in the title. The *American Encyclopedia of Education* rounded out the in-library choices.

Item 13 asked students to choose a search strategy to locate three scholarly sources. Three students who selected the correct answer on the library test answered incorrectly during the in-library test, and two students who had incorrect answers on the written test performed correctly during the in-library test. The correct answer was to search an education database for journal articles. Two of the students verbally indicated the correct response, but were unable to locate an education database. One went to the library catalog and the other chose the American Memory database. A third student indicated he had "no clue." Four other students who located the correct answer appeared to have difficulty locating an education database, as they spent a lot of time clicking around the library's Web pages. The library offers a number of links to its databases, but the researcher set the computer screen to the library's home Web page. This may not have been a database access path preferred by the students.

Overall, students' scores were fairly consistent between the two measures. Much of the variation may be accounted for by student guessing, or researcher bias in setting up the in-library test (primarily through selection of sources that may not have adequately represented item responses or setting the computer screen to unfamiliar access paths). As 78.8% of the eight in-library test items were answered consistently by the ten students, the written test appears to offer evidence that student performance on information-seeking tasks relates to their test scores.

Factor analysis

Factor analysis of test data was conducted using SPSS version 10.0 software. Bartlett's test of sphericity equaled 365.20 with a significance level of .01, and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy yielded a value of .689, which exceeds the .50 generally considered adequate for factor analysis. Minimum Eigen values were set at 1.0. Factors were extracted using Generalized Least Squares method. This method was chosen as items loaded heavily on one factor in exploratory analysis. The Generalized Least Squares method assigns weight inversely to the uniqueness of a variable, thus teasing out differences among variables. In the initial factor analysis, the researcher limited analysis to four factors with blanking set at .25. Four factors were specified to reflect the four content clusters.

The four-factor solution explained 21% of the covariance among the items. Factor One accounted for 11% of the covariance and consisted of 13 items with loadings ranging from .31 to .51. Factor One contained items from the first content cluster, but items from other content clusters were also included. Factor Two likewise contained the items from the second content cluster, with items from other content clusters present. Further review did not reveal any discernible patterns in constructs and the existence of four discrete content clusters was not confirmed.

Because the B-TILED is a new instrument, further exploratory analysis was conducted. A five-factor solution, with blanking set at .25, offered the most interpretability for the test. The five factor solution explained 23.5% of the covariance among items. When the factor solution was analyzed, Factor One accounted for 11% of

the variance and consisted of nine items with loadings from .25 to .56. Factor One contained four items from Content Cluster Two and three from Content Cluster Four. This factor was characterized as knowledge of search strategies and legal and ethical practices.

Factor Two contained three items which accounted for 3.7% of covariance. Items were from Content Clusters One, Three, and Four, and had loadings ranging from .37 to .66. No discernible pattern was uncovered for this grouping. Five items, with loadings ranging from .34 to .41, describe Factor Three. Factor Three accounted for 3.5% of the covariance. Four of the items that loaded on Factor Three were from Content Cluster One. This factor is clearly identifying, selecting, and evaluating finding tools. Factor Four contained three items with loadings from .27 to .52 and accounted for 2.9% of the covariance. Each item that loaded on Factor Four was from a different content cluster. No underlying pattern was identified to explain these item loadings. Factor Five contained four items with loadings ranging from -.34 to .37. Again, items that loaded on Factor Five were from different content clusters and no underlying pattern was found.

Factor analysis results confirm findings of other library test developers (Morner, 1993; Radcliff, 2005), who also failed to find distinct subscales of their tests. Claudia Morner (1993) offers a number of explanations for this phenomenon, suggesting that the five or six items representing each of the content clusters may be too small, that some content clusters cover too broad of a range of knowledge, and that a larger sample of students may lead to factors loading more consistently on the content clusters. This perspective would indicate administration of the test to a larger sample, additional items per content cluster, or revisions to the parameters of the content clusters.

Table 7: Factor Analysis with a Five Factor Solution

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
IT 16-B	.56				
IT 24-D	.48				
IT 21-C	.48				
IT 17-B	.46				
IT 25-D	.39				
IT 22-C	.31				
IT 14-B	.26				
IT 13-A					
IT 8-A		.66			
IT 19-C		.38			
IT 28-D		.37			-.34
IT 20-C					
IT 15-B			.41		
IT 12-A			.40		
IT 10-A			.38		
IT 9-A			.35		
IT 7-C			.34		.33
IT 23-C				.52	
IT 18-B	.25			.41	
IT 27-D	.25			.27	
IT 26-D					.37
IT 11-B					.34

Note: IT is item, A-D indicate content clusters. Factors rotated using Quartimax method.

However, the type and purpose of a test should be considered when deciding how much importance should be placed on factor analytic procedures. Clearly, tests that contain a greater number of items may better specify the construct domain and express higher coefficient values, while individual items, especially for heterogeneous constructs, have such specificity that each item may have a low correlation with the construct being measured. Test developers of library and information literacy assessment scales tend to

write one or two items for an objective that may be one of many related to a standard. A tension exists between the number of distinct objectives being measured by a test, or its breadth, and the number of items used to measure an objective, a matter of depth, that is unlikely to be easily resolved.

Finally, Thompson and Daniel (1996) report that it is not uncommon to see researchers run exploratory factor analyses and then report pattern correlations as unique dimensions of the theory. They point out the fallaciousness of logic that allows analytic methods to determine our theories, and caution that our construct definitions must be grounded in theory and not in data. Clark and Watson (1995) add that because of or despite analytic results obtained, we may wish to retain definitions that have not been empirically supported.

Reliability

According to an array of textbook definitions, a measure is considered reliable if administration of it at different times and places yields the same measurement. The goal of a test is to measure one thing – and only this thing – as precisely as possible. Two procedures promoted by developers to provide evidence of test reliability are calculating internal consistency values and investigating stability of test scores. There are two recurring issues with establishing reliability of tests. Much like the misunderstanding between construct validity and factor analysis, it appears many researchers also believe that adequate reliability can be established simply by reporting a single internal consistency coefficient.

Second, and comparable to the argument that it is inappropriate to use procedures that rely on variability of scores for demonstrating validity for criterion-referenced tests, it is improper to rely on variability based procedures for reliability evidence (Popham & Husek, 1969; Swaminathan, Hambleton, & Algina, 1974). As an example, criterion-referenced test scores can be heavily skewed, and it is not unheard of for a testing population to attain a perfect score. Although this would lead to a zero internal consistency value, Popham and Husek (1969) contend the test could still be reliable. Swaminathan et al. (1974) add that the primary evidence for establishing reliability of criterion-referenced tests is consistency of scores across repeated administration.

Both stability and internal consistency procedures were conducted. Stability of the instrument was measured by a test-retest procedure whereby the written test was administered twice, over an approximate two week interval. To measure internal consistency, data were submitted to K-R 20 calculation.

Stability

Stability was assessed by comparing test scores from the written test with a later administration of the same test. Eleven students, who indicated their willingness to participate in further testing after the first administration, took the test a second time. Students were given the same written test form and instructions as they received earlier. Approximately two weeks lapsed between test administrations.

Table 8 summarizes results of the eleven participants. Of the eleven pairs of 22 items, or 232 pairs of items for the test and retest, 172 pairs matched across test

administrations. With eleven participants, the mean change was 2.4 items out of 22; therefore, 74% of items matched from one test administration to the next. The test-retest results indicated general stability over time. As noted previously, this procedure has been recommended as constituting primary evidence for criterion-referenced test reliability (Swaminathan et al., 1974). Repeated administrations should confirm or negate these preliminary results.

Table 8: Test/Retest Stability Results

Initial Test	Score	Retest	Score	Change	Matched Pairs
1A	13	1B	11	-2	14
2A	18	2B	13	-5	17
3A	12	3B	17	+5	15
4A	10	4B	8	-2	12
5A	12	5B	14	+2	16
6A	8	6B	8	0	13
7A	9	7B	13	+4	12
8A	13	8B	13	0	15
9A	18	9B	17	-1	19
10A	11	10B	14	+3	19
11A	19	11B	17	-2	20
Total				26	172
Mean	13		13.2	2.4	

Internal Consistency

Internal consistency coefficients provide information regarding unidimensionality and measurement error. These coefficients are affected by several characteristics of the test, including intercorrelations among the items, length of the test, and test content. Higher reliability values reflect that test items tend to cohere, or that test takers who answer a given question correctly are more likely to correctly answer other questions measuring the same construct. Lower internal consistency values can indicate problems with theory, test development, the testing situation, or student heterogeneity (Clark & Watson, 1995; Swaminathan et al., 1974).

Low values due to theory occur when the theory is inadequate and test content and item development pull from poorly defined constructs. Similarly, test development procedures that negatively influence internal consistency estimates may be traced to poorly worded or presented items, or too few items. A well known property of educational and psychological tests is that the longer these tests are, the higher the internal consistency scores they yield. This is attributed to true score variance increasing more rapidly than error variance as test length increases (Ebel, 1972; Gardner, 1970). In a recent information literacy study, Cameron (2004) added 14 items to a test originally containing 33 items, with a resulting increase in coefficient values from .69 to .77. Kehoe (1995) reports that values as low as .50 have been considered satisfactory for tests with as few as 10 or 15 items, with the consequent expectation that tests with over 50 items should yield coefficient values of .80 or higher.

Test administration characteristics must also be considered. Researchers should ask if the sample was nonrepresentative in some way or, to answer the question of student heterogeneity, how likely is it that the construct is present in the sample. If it is not known whether test takers possess a sufficient degree of the construct, then the internal consistency coefficient reveals little information. Certainly, items are less likely to cohere if test takers have a low level of the construct. Regardless of the type of test, each explanation must be considered before interpreting the internal consistency coefficient and taking steps to enhance it.

Test results. The K-R 20 coefficient was .675 for the test. Developers of psychological tests that are used to make important decisions have traditionally looked for reliability estimates in the area of .90. In their investigation of psychometric procedures reported in the literature, Clark and Watson (1995) point out that there are no longer any clear standards regarding what internal consistency coefficient levels are considered acceptable. They note that minimum standards of .80 and .90 are recommended for basic and applied research, but it is not uncommon for researchers to consider values in the .60 to .70 range. Information literacy test development studies are also inconsistent in their agreement of acceptable values. For example, Cameron (2004) reports a .69 value as "quite adequate for program evaluation purposes" (p. 211) for a 33 item cognitive skills test, while Gratch Lindauer and Brown (2004) suggest that a coefficient of .76 is not sufficiently high for a test with similar purpose and construction.

A number of explanations may be offered for the modest internal consistency value returned for the test. Among these explanations are test length, the heterogeneous

nature of the construct (and the attempt to measure the range of it with one scale), and the question of whether the construct is present in the sample tested. One procedure often followed to enhance internal consistency values entails eliminating lower-cohering items. A review of the inter-item correlations indicated a higher value could be achieved if two of the items were removed. Although prevailing practice is to withdraw the items from the test to raise the value, Clark and Watson (1995), in an articulation of Loevinger's "attenuation paradox," proffer a more considered approach. They argue that often approaches to internal consistency occur at the expense of breadth, and that retaining those items which correlate most highly is redundant. While this procedure serves to increase internal consistency values, it also may result in an overly narrow scale that does not assess the construct optimally and which could compromise validity.

Popham and Husek (1969) note that high inter-item correlations or high test-retest correlations *can* be used as evidence of reliability. However, as with all criterion-referenced test validation procedures based on variability, strong or positive results can support claims, but low or negative results do not negate a test's reliability. Without doubt, a brief test that attempts to measure such a diverse construct as information literacy among a heterogeneous population of test takers is not going to result in an exceptionally high internal consistency coefficient.

Content cluster results. Internal consistency statistics for each of the four content clusters were calculated and are displayed in Table 9. K-R 20 values for content cluster A, identifying, evaluating, and selecting finding tools, content cluster B, demonstrating knowledge of searching techniques, content cluster C, evaluating and selecting sources,

and content cluster D, knowledge of legal and ethical practices, were .450, .433, .334, and .174 respectively. Internal consistency statistics for the content clusters ranged from moderate to low. This may be attributed to the same characteristics mentioned for overall test results.

Table 9: Internal Consistency of the Four Content Clusters

Content Cluster	Mean	SD	Number of Items	K-R 20
A	2.63	1.34	5	.450
B	3.39	1.48	6	.433
C	2.91	1.42	6	.334
D	3.04	1.06	5	.174

Analysis of Test Data by Respondents' Characteristics

In addition to the 22 content items, 13 demographic and self-percept questions were included in the test. The demographic questions asked for information regarding gender, ethnicity, student classification, and length of enrollment at the university. Two questions asked students to self-rate their ability to search library databases and the Internet, and four questions were dedicated to ascertaining students' exposure to library instruction. These questions were asked in an effort to determine if a link exists between test scores and the demographic or self-percept variables. Cross tabulations for the

variables of gender, ethnicity, student classification, and length of enrollment with mean score were calculated. Self-rated library searching ability, Internet searching ability, and intensity of exposure to library instruction with mean score were also analyzed.

Gender

A cross tabulation of gender with mean score did not reveal any significant differences. Of the 170 respondents who answered the question, females comprised 80.00% of the sample and males 20.00%. With 136 responses, the mean score for females was 12.05 (SD=3.75), with scores ranging from 3 to 20. The mean score for the 34 males was 11.44 (SD=3.69), and scores ranged from 2 to 19. No statistically significant differences were found between groups; however, as a group females scored .61 of a point higher than males. Future test administrators should continue to monitor results to see if the test is equally reliable in measuring the construct of information literacy of males and female. Table 10 presents a summary of the breakdown by gender.

Table 10: Mean Scores by Gender

Gender	Mean	SD	Number	Valid Percent
Female	12.05	3.75	136	80.00
Male	11.44	3.69	34	20.00

Note: N=170

Ethnicity

Ethnicity compared to mean score likewise did not reveal any statistically significant differences. Of the 166 students who answered the question, the 137 students who identified themselves as White or European-American had a mean score of 12.18 (SD=3.73), the 13 students who indicated Hispanic or Latino had a mean score of 10.92 (SD=3.75), the 12 Black or African-American in origin students had a mean score of 10.77 (SD=3.42), and the 4 Asian or Asian-American in origin students had a mean score of 10.75 (SD=5.62). Summary data are offered in Table 11. Although no statistically significant differences among groups were found at the .05 level for ethnicity, White or European-American students scored over a point higher on the test when compared to other ethnicities. Again, results of future test administrations should be examined to see if the test adequately discriminates knowledge of information literacy among ethnicity.

Table 11: Mean Scores by Ethnicity

Ethnicity	Mean	SD	Number	Valid Percent
White or European-American	12.18	3.73	137	82.53
Hispanic or Latino	10.92	3.75	13	7.83
Black or African-American	10.77	3.42	12	7.23
Asian or Asian-American	10.75	5.62	4	2.41

Note: N=166. The "other" category, containing one response, is not represented here.

Student Classification

The test was designed specifically for undergraduate students enrolled in a teacher education program, so student classification responses were limited to freshman, sophomore, junior, and senior. The relatively small number of freshmen and sophomores was not surprising, as students are generally accepted into the program after completion of their general education requirements. Of the students who answered the question, freshmen comprised 8.00%, or 12, of the 150 responses, sophomores 6.77%, or 10, juniors 32%, or 48, and seniors 53.3%, or 80.

The mean score for freshmen was 10.42 (SD=2.75), with the number of correct scores ranging from 7 to 15. The mean score for sophomores was 11.50 (SD=3.60), with a range in scores from 6 to 18. With a mean average of 10.38 (SD=3.27), juniors were slightly lower than sophomores and fairly equal to freshmen. The range in correct scores for the 48 juniors was 4 to 18, which was greater then freshmen or sophomores. Seniors were the largest group to answer the test, and with 12.55 (SD=3.93), also had the highest mean score. Correct answers for seniors ranged from 2 to 20. Higher mean scores for seniors may be attributed to continuing exposure to relevant instruction or to student maturation. A summary of statistics is offered in Table 12.

Table 12: Mean Scores by Student Classification

Student Classification	Mean	SD	Number	Valid Percent
Freshman	10.42	2.75	12	8.00
Sophomore	11.50	3.60	10	6.77
Junior	10.38	3.27	48	32.00
Senior	12.55	3.93	80	53.33

Note:N=150

Length of Enrollment

Students were also asked the length of time they had been continuously enrolled at the institution. Enrollment was cross tabulated with scores and revealed increasing mean scores on the test the longer the student had been enrolled. The 41 students who had been enrolled for less than one year had mean scores of 10.71 (SD=3.49), compared to 11.51 (SD=3.49) for the 55 students who indicated they had been continuously enrolled for 1 to 2 years, 12.26 (SD=3.78) for the 50 students enrolled from 3 to 4 years, and 14.26 (SD=3.73) for the 23 students who were continuously enrolled for more than 4 years. Table 13 summarizes test scores by length of institutional enrollment.

Table 13: Mean Scores by Length of Enrollment

Length of Enrollment	Mean	SD	Number	Valid Percent
Under 1 year	10.71	3.49	41	24.26
1 to 2 years	11.51	3.49	55	32.54
2 to 4 years	12.26	3.78	50	29.59
Over 4 years	14.26	3.73	23	13.61

Note: N=169

Level of Instruction

Four questions on the test asked to what extent the student had received library instruction. These questions sought to determine if a link exists between the amount of library instruction received and scores on the test. Level of exposure to library instruction was determined by calculating the number of positive responses to the four questions. For example, if a student answered "no" to all four instruction questions, they were assigned an exposure level of 'none." Similarly, a positive response to one of the four questions resulted in assignment to the "minimal" category, a positive response to two of the four questions was considered "moderate," a positive response to three of the four questions was considered "high," and a positive response to all questions was considered "intensive."

Mean scores were compared to instruction levels and are presented in Table 14. The 42 students who responded that they had no prior exposure to library instruction scored close to the test average (M=11.76, SD=3.73), while the 32 students with minimal

exposure averaged more than a point higher (M=12.84, SD=3.32). The 47 students with moderate exposure (M=12.38, SD=3.61) and the 36 with high exposure (M=12.11, SD=3.69) demonstrated a slight inverse relationship in score to level of library instruction. As students' library instruction exposure increased, test scores decreased. This was most apparent in students with intensive exposure to library instruction. The 15 students reporting all four instructional contacts, with a mean test score of 9.07 (SD=4.11), scored markedly lower than the four other groups.

Table 14: Mean Scores by Level of Instruction

Level of Instruction	Mean	SD	Number	Percent
No Instruction	11.76	3.73	42	24.42
Minimal	12.84	3.32	32	18.60
Moderate	12.38	3.61	47	27.33
High	12.11	3.69	36	20.93
Intensive	9.07	4.11	15	8.72

Note: N=172

This finding appears perplexing, but is not unprecedented. Tunon (1999) also found that increased exposure to library instruction of doctoral students in education did not translate into significantly better dissertation literature reviews. Kunkel, Weaver, and Cook (1996) offer another explanation. In study that assessed undergraduate students' library skill levels, they reported that it is not the number or frequency of library instruction sessions that best predicts test scores, but the frequency with which students

receive assignments requiring library research. For the current study, it is difficult to draw conclusions from these results as the number and frequency of library-related assignments was not gathered, there were only 15 students in the "intensive" category, and the standard deviation of 4.11 was higher than standard deviations of other categories.

Library Database and Internet Searching Ability

A number of researchers (cf., Fox & Weston, 1993; Greer et al., 1991; Maughan, 2001) have reported that students tend to overestimate their searching abilities, so two questions were posed that asked students to self-rate their library and Internet searching skills. Students selected from responses of "excellent," "good," "average," and "poor." Mean scores on the test were compared to students' self-percepts of library database and Internet searching ability. Library searching ability summary data are presented in Table 15.

Consistent with the literature, students who were most confident in their ability to search library databases scored lower than students who reported in the "good" or "average" range. However, students who considered their library database searching skills as "poor" tended to score the lowest on the test. Fifty, or 29.07% of students, rated their skills as excellent. Students who scored their abilities highest had a mean average of 11.60 (SD=3.90) on the test, with a range of scores from 2 to 20, which is the complete range of scores available. The 80 students who rated their library skills as "good" averaged 12.25 (SD=3.77) on the test, with a range in scores from 3 to 19. Mean scores

for "average" were 12.08 (SD=3.48), and the 37 students who assessed themselves at this level had scores that ranged from 5 to 19. The remaining five students who rated their library skills as "poor" had a mean score of 10.4 (SD=3.78), and a range in scores from 6 to 16.

Table 15: Mean Scores by Library Database Searching Ability

Self-Rated Library Ability	Mean	SD	Number	Percent
Excellent	11.60	3.90	50	29.07
Good	12.25	3.77	80	46.51
Average	12.08	3.48	37	21.51
Poor	10.40	3.78	5	2.91

Note: N=172

Students who rated their Internet searching abilities as "excellent" tended to score higher than students who had rated their library searching skills at the same level. The 93 students who rated their Internet searching skills the highest had a mean score of 12.26 (SD=3.73) on the test, and a range in scores from 2 to 20. Students who rated their skills as "good" (n=61) had a mean score of 11.34 (SD=3.77), and a range of 3 to 19. The 17 students who rated their Internet skills as "average" had a mean score of 12.24 (SD=3.25) and a range in scores from 5 to 18. Only one student reported Internet searching skills as low, and that student scored a 19 on the test. These results support earlier findings that students tend to overestimate their library searching ability; however, patterns in self

report of Internet searching skills are not so easy to discern. Those who rated themselves as "excellent" searchers did have the highest score on average, but "average" ability students scored almost as high while those who rated themselves as "good" scored almost a point lower on a 22 point scale. Most unexpected was the sole student who rated her skills as "poor," yet scored very highly on the test. Internet searching ability comparisons are located in Table 16.

Table 16: Mean Scores by Internet Searching Ability

Self-Rated Internet Ability	Mean	SD	Number	Percent
Excellent	12.26	3.73	93	54.07
Good	11.34	3.77	61	35.47
Average	12.24	3.25	17	9.88
Poor	19.00	0	1	.58

Note: N=172

Overall, analysis of demographic variables did not indicate any sizeable differences among student categories. However, some variation in test scores among groups of respondents was found, which suggests that the test should continue to be examined for its ability to effectively measure the information literacy skills levels of participants belonging to subgroups of education students.

Passing Score Calculation

The usefulness of the cut score should be considered in relation to the purpose to which the B-TILED will be used. If the test is used to determine mastery of the content and an individual student score is needed to reflect mastery, then calculating a passing score may be useful as an overall indication of students' skills and knowledge levels. However, if the researcher wants to determine the level of students' knowledge in relation to discrete subscales or content clusters, then utility of the cut score is limited. Passing score determinations are designed to measure one construct, and the calculation treats each item the same. It is possible that a student could answer every question in content cluster A, B, and C correctly, but miss every item in content cluster D. The calculation may miss a unique set of knowledge and skills that is essential to be considered information literate.

Researchers (cf., Berk, 1986; Lord, 1974; Nitko, 1970; Simon, 1969; Skager, 1974) have reminded us that the cut score is an artificial means of establishing a hypothetical boundary on continuous data. As such, they advise that when individual decisions are not required, as in the case of program evaluations, continuous distribution of scores is preferred. In these cases an examination of cohort test scores can be used for gross analysis and a closer look at subscales may ensure that the competency is being addressed during instruction.

The passing score for the B-TILED was calculated using variants of the Angoff method. The procedure consisted of several steps. First, a panel of five experts examined each item and estimated the probability that the "minimally acceptable" person

would answer the item correctly (more specifically, judges were asked what proportion of one hundred information literate test takers should answer each item correctly). The sum of the panel's estimated proportions was averaged to arrive at a preliminary passing score. This data is presented in Table 23, in the Initial Reviewer Rating column. The preliminary passing score calculation revealed 55.5% of items would need to be answered correctly for the test taker to demonstrate acceptable levels of information literacy skills knowledge as determined by the panel.

Berk (1986) adds that the cut score setting method should take into account the results of the field test and set standards in a realistic range. Therefore, after test administration, judges reviewed item difficulty levels and test score distributions and frequencies and were allowed to revise their original estimates. Although the mean test score for the field administration was 54.42%, the judges' adjusted passing score level rose to 58.8%. The judges' adjusted scores were influenced by item difficulty levels and the fact that students who completed the test had varying levels of library instruction; the sample included students at various stages of their program. Judges commented that they would expect higher scores from a group of instructed students. Item difficulty levels and adjusted reviewer ratings are also presented in Table 17.

Finally, each item estimate was further adjusted by the standard error of the mean. As noted earlier, measurement error can be introduced in any number of ways. Adjusting the standard error of the mean is one method used to account for measurement error. This error can allow for false positive scores, where an individual may not be minimally competent yet still manage a passing score, or false negative scores, where an individual does not achieve a passing score yet is competent.

Table 17: Cut Score Statistics

Item Number	Judges' Initial Scores	Item Difficulty	Judges' Adjusted Scores	SEM Adjusted Score
7	.55	.49	.55	.537
8	.55	.32	.50	.487
9	.50	.57	.60	.587
10	.45	.41	.45	.437
11	.40	.39	.45	.437
12	.70	.68	.70	.687
13	.65	.65	.70	.687
14	.65	.60	.65	.637
15	.65	.42	.60	.587
16	.60	.59	.60	.587
17	.65	.73	.75	.737
18	.45	.65	.60	.587
19	.55	.36	.45	.437
20	.45	.43	.45	.437
21	.45	.69	.70	.687
22	.35	.57	.60	.587
23	.40	.42	.40	.387
24	.60	.57	.60	.587
25	.85	.89	.90	.887
26	.20	.34	.35	.337
27	.80	.81	.80	.787
28	.45	.42	.45	.437

For high stakes tests, where assigning competency is critical, such as medical competency exams, measurement error is generally *added* to the score to minimize false

positives. For lower stakes tests, where admission to the next level is not so highly monitored, measurement error can be *subtracted* from the scores. This gives the individual the benefit of the doubt in relation to his or her test score. Individual item percentages were adjusted down to allow for test error measurement and to minimize false negative scores. Based on these calculations, test takers needed to achieve a score of 57.5% to be considered acceptably competent. Of the 172 students who completed the B-TILED, 76 met that goal.

Ultimately, the most significant factor for establishing a cut score is the importance of the decisions to be made with the results. High stakes uses include assigning grades to students and making placement decisions based on scores. At this point, it is highly unlikely information literacy testing will be used to inform these decisions. While it may be of interest to assign an individual score, the cut score calculation may be of limited utility for most information literacy test developers as it is anticipated assessment efforts will be used more for program evaluation purposes than passing a student based on the test score.

Summary

Results of procedures conducted to validate the B-TILED were reported in this chapter. The test was administered electronically and in print to a total of 172 students enrolled in an undergraduate teacher education program. Analysis and report of results included descriptive statistics of the sample and the test, and the two administration modes. Item level analysis results reported data on item difficulty levels, discrimination

indices, and the percent of respondents choosing each response for each test stem. Results of various procedures, including content development, criterion-related validity, factor analysis, stability, and internal consistency, were offered as evidence of validity and reliability. Analysis of test respondents' characteristics was also reported, as were results of the passing score calculation. Overall, the combined evidence supports validity claims of the B-TILED. In Chapter Five, these results are further discussed, as are limitations, recommendations for future research, and practical and theoretical significance of the study.

CHAPTER FIVE
CONCLUSION

The goal of this study was to develop a valid, easily administered instrument that assesses education students' information literacy skills levels. At the onset, this goal seemed very attainable. After all, test validation procedures described in research methodology texts and reported in many studies were very straightforward. However, it quickly became apparent that, like so many others, this researcher had a very superficial understanding of how tests are developed and validated. Clark and Watson's (1995) critique of test validation methodologies confirmed this suspicion, and motivated me to abandon the studies and methodologies reported in the literature and instead consult primary texts on the topic.

The discussion among early test development theoreticians focused on what procedures constitute the best evidence for validity of criterion-referenced tests. More specifically, these researchers argued for procedures that did not rely on variability in scores. The themes put forward in these early works were never refuted; instead, they appear to have grown unfashionable or perhaps simply forgotten. More modern approaches to test validation, even for criterion-referenced tests, place considerable emphasis on internal consistency values and factor analytic results. Deciding how to balance these divergent perspectives presented a quandary that was not easily resolved.

With moderate internal consistency values and a factor analytic pattern that did not support original expectations, the B-TILED would not be deemed trustworthy by those who hold fast to the modern perspective. Conversely, not acknowledging the centrality most researchers place on "reliability" scores and factor structures could be

perceived as naïve, and serious thought not accorded to the study's results. After considerable deliberation, this researcher decided the most appropriate way to proceed was to describe in detail the study's methods and results and bring the discussion of testing issues to the table, allowing the reader to judge the relative merit of the argument presented and how it affected the study's methods, interpretation of the results, and ultimately the validity of the test.

The purpose of this disclosure is not to be overly modest, but to acknowledge that two different perspectives exist regarding what constitutes acceptable evidence for test validation. The logic presented by early analysts was more compelling to this researcher than more modern approaches and consequently had impact on the current study. The reader can be reassured the study was conducted under the watchful eye of both Project SAILS administrators and the dissertation committee. For those interested, the researcher's qualifications are also offered for review (see Appendix K).

Based on results of criterion-referenced test validation procedures endorsed by early researchers, the study goal was satisfactorily attained. The chapter's conclusions and discussion are presented around the study's themes, and begin with a description of the research problem, which was the need for a test such as the B-TILED, and results of the study. The discussion then addresses practical use of the instrument and theoretical significance of the study, and revisits test validation procedures. Limitations of the study and suggestions for future development are also included. The chapter ends with a brief summary.

Summary of Research Problem

Information available both in print and on the Internet ranges widely in scholarliness and quality, yet electronic dissemination is generally preferred due to the freedom from temporal and geographic restrictions (UCF Libraries, 2005). Initial reports indicate that students increasingly rely on information gleaned from the Internet, yet are not consistently critical in their evaluation of the material. The need for information competent individuals who can effectively contribute to today's workforce has been recognized by many professional and accrediting bodies. It is this attention, combined with appeals from academics and employers, which has led to recent government and industry-supported initiatives to develop information literacy tests.

A review of existing information literacy and fluency assessment instruments revealed a need still existed for an instrument such as the B-TILED. The Project SAILS information literacy test is designed to test information literacy skills at the general education level. With the additional emphasis on technology skills, the ICT Literacy Assessment evaluates a broader range of skills than the SAILS test, yet also assesses skills at the general education level. Other nationally recognized instruments, such as the information skills test developed by James Madison University and used by several other institutions, assess skills typically found at the general education level.

The push for credible information literacy tests has not been limited to general assessments. Recognizing that information production and dissemination is unique to the academic discipline, ACRL (2000) placed a call for the development of instruments that assess discipline-specific information literacy knowledge and skills. With backing from

IMLS, Project SAILS facilitated the call by offering fellowships to attract test item writers for various disciplines. This burgeoning national level attention, supported by numerous federal and academic institutions and private and professional organizations, indicates the level of interest in and need for continued development of both general and discipline-specific information literacy skills tests.

Practical Significance

To date, results of instruction program evaluation studies have contributed little to our understanding of instructional efficacy. It is perhaps the lack of rigorously reviewed instruments that has most prohibited systematic investigation of the topic. Credible information literacy assessment instruments, such as the B-TILED, are needed for a variety of institutional purposes. Additionally, objective instruments are most effectively used in conjunction with other measures of information literacy skills as part of a consistent and on-going assessment program.

Test results can be used to inform internal decision-making, to provide evidence for external purposes, and to indicate individual mastery of content. Internal uses of test results include evaluating and improving the instructional program, increasing institutional support, and justifying the instructional program. Examples of external uses of the data are benchmarking with similar institutions and providing outcome measures for accreditation purposes.

For many, information literacy assessment may be limited to providing data for program review and accreditation evidence. In these cases tests that rely on cohort

analysis and compare scores at the institutional level may suffice. However, there is question as to whether education students are conceptually prepared to take full advantage of the information available to them. Testing may be one method used to identify individual student skills. For this reason criterion-referenced tests with established passing scores may be useful. The B-TILED can be used across purposes, from internal review of instructional efficacy to external benchmarking comparisons to assessing individual mastery. The following illustration offers one example of how the B-TILED can be used in an institutional assessment program.

American Association for Higher Education [AAHE] assessment director Peggy Maki (2002) admonished institutions of higher education for their ad hoc, "episodic," and accreditation-driven assessment efforts. Instead, Maki suggests that assessment efforts driven by institutional curiousity and sustained over time yield richer and more informative data. Maki identifies key components for developing a successful assessment plan. These include: identifying student learning outcomes, providing sufficient educational opportunities for learning, deciding what types of assessments to use and when to use them, and collecting data, interpreting results, and deciding if and how the results will effect change.

Considerable progress has been made toward meeting the first key component of an assessment plan, identifying student learning outcomes. ACRL (2000) has developed standards, performance indicators, learning outcomes, and objectives for a broad range of information literacy skills that are widely recognized and almost universally accepted. The second component, providing sufficient educational opportunities for learning, varies according to the institution, its resources, and its commitment to information literacy.

Instructional opportunities can consist of reference desk transactions, one-shot instructional sessions at the library, for-credit courses, systematically sequenced instruction, and be face-to-face or distance, synchronous or asynchronous, and librarian or course faculty facilitated. However, "sufficient educational opportunities" also includes occasions to practice and apply these skills. Faculty must assign substantive term papers and projects that require students to apply information literacy skills and be prepared to provide continuous and on-going feedback.

The third key component of an assessment plan, deciding what types of assessments to use and when to use them, is the most critical. Reasons for assessing must be determined prior to deciding which assessments to use and at what points they should be administered. For example, faculty may assume students entering a program have gained information literacy skills elsewhere, so assessments of entering students may identify deficiencies that can be targeted with instruction. Assessment at this juncture also provides baseline data for comparison to later performance. Next, to investigate the efficacy of a particular method of instruction or to compare instructed students with those who have not received instruction, assessments can be conducted at various points throughout a program. Results of these assessments can offer valuable formative information to the student as well as inform instructional decisions. Finally, summative evaluation can provide individual performance scores as well as document whether skills have developed over time when compared to baseline data.

These purposes for assessment occur at both the general education level and the disciplinary level, and objective tests are available for each level. At the general education level, the ICT Literacy Assessment and SAILS tests are examples of

instruments that are used to assess students' information literacy skill levels. Transitioning to the major, discipline-specific measures are indicated. The B-TILED is one test developed for use with education majors.

If student learning outcomes are identified and sequenced throughout the program, then the same test can be administered repeatedly, with the expectation that scores would increase as students progressed through the program (Maki, 2002; Skager 1974). This holds true at both the general education and disciplinary level. Given this, the ICT Literacy Assessment or SAILS tests could be administered upon entry to the institution, throughout the general education program, and again prior to transitioning to the major. Similarly, the B-TILED could be administered at the beginning and end of the program, and at points in between. If the tests are administered at varying times during a student's matriculation, then data regarding where the student is in the program should be gathered in addition to the student's classification.

The following chart provides one example of a transitioning assessment plan. The chart is limited to using objective tests to measure cognitive knowledge. A fourth information literacy assessment point, as yet relatively uninvestigated, is post completion of the program or after graduation. This point is not included in the chart as it is more likely employer or alumni surveys or other assessments would be conducted. The remaining key components of an assessment program, collecting data, interpreting results, and deciding if and how the results will effect change, are self-explanatory and flow from the first three steps.

Purpose	Stage	Test
General Education Level		
Collect baseline data, Identify deficiencies to target instruction	Entry to institution	ICT Literacy Assessment or SAILS test
Instructional efficacy, Formative student feedback	Points throughout program	ICT Literacy Assessment or SAILS test
Institutional effectiveness, Summative student scores	Completion of general education requirements	ICT Literacy Assessment or SAILS test
Major or Program Level		
Collect baseline data, Identify deficiencies to target instruction	Entry to major	B-TILED
Instructional efficacy, Formative student feedback	Points throughout program	B-TILED
Institutional effectiveness, Summative student scores	Completion of coursework or after clinical experience	B-TILED

This plan only describes assessment in the cognitive domain using objective testing instruments. Yet, more must be known regarding the impact of information literacy instructional programs on students' attitudes, beliefs, feelings, and behaviors. Evaluation of information literacy instruction should not be limited to assessing its impact on the cognitive domain, as content knowledge and self-efficacy are both important in preparing information literate students (Kuhlthau, 1993; Mensching, 1987; Nahl-Jakobovits & Jakobovits, 1993; Ren, 2000). The ability to apply cognitive

knowledge of the construct, as exhibited by relevant behaviors of information competent students, is also needed. To investigate whether information literacy instruction impacts student actions related to information seeking the B-TILED should be used in conjunction with affective and behavioral assessments. Assessment results can also be considered in conjunction with other institutional-level data. For example, results can be used to compare performance on NSSE indicators with benchmark institutions. If the number and length of assigned papers is relatively low, then this provides additional insight into student opportunities to practice and learn instructional goals.

Ultimately, no single measure can capture the complexity of learning. The justification for a brief, selected-response test such as the B-TILED is the need for a method that is easy to administer and produces readily analyzable data; the qualification is that multiple forms of assessment are needed to truly gauge student performance and program effectiveness. The B-TILED, therefore, is offered as one tool in a repertoire of information literacy assessment instruments. Multiple methods of assessment, administered at critical points throughout the learning process, are necessary to validate an assessment program and successfully measure the range of student achievement (Association of American Colleges and Universities [AACU], 2005; NCATE, 2002). The AAHE (2005) writes that learning is multidimensional, integrated, and revealed in performance over time. Assessment should be, as well.

Theoretical Significance

Credible instruments are needed for a variety of institutional purposes, but perhaps even more imperative than documenting positive instructional impact for program review is the need to gain a better understanding of information literacy and its viability as a theory. Theory is fundamental to systematic inquiry, and consistent relationships between various theoretical concepts are established, verified, and extended through continuous testing (Clark & Watson, 1995; Kerlinger & Lee, 2000; Shulman, 1999).

Information literacy is a concept that has come into existence in the last 30 years, and use of the term and what it encompasses has dramatically expanded during this time. Information literacy can be conceived of as a construct that attempts to explain the relationship between efficient, effective, and ethical use of information combined with a critical understanding of how information is produced, disseminated, and organized and its relation to the information seeking process. Although information literacy as a theoretical construct is appealing, claims as to its validity remain unconfirmed. More information, gathered via systematic inquiry, is needed to build the theory. Questions abound that call for answers, and consist of:

- What does information literacy include? What are its boundaries? What is unique? Do test results correlate with results of critical thinking and problem solving tests, and tests of technology skills? Does this support the validity of information literacy as a distinct theory?

- Is information competence predictive of academic success? Personal success? Success in the workplace? Is presence or absence of the construct expressed in observable behavior? If so, how?

- Are information literacy skills correlated to self-efficacy or library anxiety? How does this explain student persistence or success? If relationships are found, how does it impact instructional and curricular planning?

- Do test results indicate whether the B-TILED is consistent with the theory of the construct? How the theory was operationalized? Are there other criteria or methods that could be used to improve measurement of information literacy skills of education students?

As observations accrue hypotheses are tested and theory is refined or further elaborated. The process is iterative in the sense that the scale is informed by the theory of information literacy, yet scale results can contribute to elaboration of the theory. It is generative in the sense that progress toward understanding theory is cumulative and continually builds on existing evidence. The B-TILED can be used to collect data and theory build, in turn leading to a deeper understanding of the construct of information literacy.

Test Validation Issues

A discussion of test validation procedures was not original to the design of the study, but conflicting perspectives and methodological critiques were uncovered during the process. These discussions centered around three issues and were comprised of the complexity of determining scale reliability and validity, the use of procedures that rely on variability of scores for validating criterion-referenced tests, and the calculation of individual passing scores.

Test validation evidence widely presented in the literature could lead one to conclude that construct validity and reliability can be established simply by respectively

reporting factor analytic results and internal consistency values. Regardless of the test's purpose, establishing reliability and validity of a scale requires much more evidence than what is reported from a single procedure. A body of evidence, collected from multiple procedures, is generally offered to demonstrate adequate scale validity and reliability.

Second, the utility of procedures that rely on variability in scores for validating criterion-referenced tests was discussed. Early researchers regard these procedures secondary, and do not consider them essential as commonly indicated by more recent reports. These procedures, comprised of factor analysis and internal consistency calculations, happen to be those most relied upon for psychological tests. This led to wide reporting of results in the literature and an unquestioned acceptance of their use in establishing test validity. It is the unquestioned acceptance that is of most concern, as results of these procedures may provide evidence, but do not necessarily constitute the most important evidence.

Finally, a significant portion of the paper was devoted to the case for calculating cut scores and an articulation of one procedure suggested for establishing them. Berk (1986) suggests that researchers should rely primarily on judgmental procedures for suggesting the initial cut score. He further recommends that these preliminary scores be confirmed or adjusted using empirical procedures. The B-TILED cut score was calculated using variations of the Angoff method. It is prudent at this point to remind test developers and B-TILED administrators that the calculation is, indeed, an estimated score. Confirmation of the score is needed and the Harris procedure, described in the subsequent Suggestions for Future Development section, is recommended.

Limitations of the Study

Several limitations are associated with the study. Some limitations were oversight and others occurred as a result of logistical constraints. These limitations are addressed to understand implications for the study's results and to inform revisions to the B-TILED or future test developers. Test design and population sampling and the need for on-going construct validation comprise the bulk of the concerns. A discussion of each of these areas follows.

Test Design Limitations

The purpose of asking demographic and attitudinal questions is to allow for analysis of test results by respondents' characteristics. If the instrument is credible, and has been proven to measure the construct rather than some other attribute, then results of the analysis can reveal important information regarding particular subgroups of the population being studied. For example, at the most basic level, the researcher would want to know if the student completing the test was instructed and to what extent the student was exposed to various instructional opportunities. Researchers may also want to know if instruction has equal impact on males and females, on freshmen through seniors, etc. For purposes of test validation, it is essential test developers ascertain the extent the test equally assesses skills levels across groups. Although student demographic and attitudinal questions can be customized to reflect local interest in areas of inquiry at the point of test administration, at the test validation phase all possible subgroups should be identified.

Unfortunately, no questions were included on the scale relating to respondents' native language. This leads to the question of whether the test can be used to assess non-native English speaking students. One item included on the test asks about ethnicity, but responses group both native and foreign-born individuals into the same category. Although Scantron answer sheets limit users to five responses, an additional question could have been easily added that asked if the test taker was a native English speaker. This level of granularity provides researchers more information regarding the effect of their instructional programs for non-native English speakers. Test scores among ethnicities were not statistically significantly different, but they did vary. Despite initial item review by two non-natives for language clarity, it is entirely possible this variation in scores was due to language constraints rather than ethnicity.

Researchers may also want to consider the full range of instructional opportunities available to students. For this study only four instructional opportunities were included and they did not allow for all types of instruction and instructional modes available to students. For example, the institution offers a variety of online tutorials, streaming videos, and reference services that were not listed on the scale. If researchers are interested in how much and what type of instruction students have completed and its possible impact on performance, then care must be taken to include all instructional modes, otherwise explainability is naturally constrained.

Another student characteristic researchers may wish to investigate in relation to information literacy skills is the frequency of assignments requiring library use. Initial reports (Kunkel et al., 1996) suggest that educational opportunities that consist of assignments that require library research are better predictors of skills attainment than

frequency of library instruction sessions. Researchers need to continue to collect data to monitor this finding, which has potential to emphasize the importance of the teaching faculty in the acquisition of students' information competence skills.

In sum, a more critical initial review of non-content items may have allowed for additional useful information for analyzing student characteristics. Suggestions for improvements were made in two areas, language and instruction. Both of these suggestions were based on observations from colleagues. For future developers, it is recommended that another step is added to the procedures. Similar to the role that content experts played in assessing content questions, review of demographic and other self-percept items by key people at the institution may assist in identifying all relevant characteristics.

Sampling Limitations

Perhaps the most serious study weakness lies with the selected population and the procedure used to identify the sample. The goal of the study was to develop an education-specific information literacy cognitive skills test that could be used across institutions. However, only students from only one institution were selected as participants. Although the population is relatively diverse in terms of ethnicity and student academic classifications, it is highly unlikely that students from one institution represent the range of characteristics found nationally. Certainly, school size and regional differences exist, and future administrations of the test may confirm or disprove

expectations that the test is equally adequate for assessing students from varying sizes of institutions, libraries, and programs, and differing regions of the country.

Sampling that occurred within the selected institution was also not without concern. As described in the study's Methods, numerous and recurring challenges arose that were connected to inclement weather, nonworking test links, and wrongly identified student populations. The study also relied on self-reported demographic information. Project SAILS administrators compared student self-reported instruction levels with institutionally maintained records and found that students' answers varied dramatically from information taken from class lists (Radcliff, 2005). This led Project SAILS to recommend that test administrators either directly pull data from institutional records or confirm student-supplied answers with institutional data. The current study did not satisfy this recommendation as instruction records are not maintained to the level of student identification information.

On-going Construct Validation

A brief, multiple-choice test has limits for assessing the complete range of a complex set of knowledge and skills. Two procedures, factor analysis and internal consistency, reveal results that are markedly lower than what many researchers consider adequately acceptable. Given current emphasis on these scores, these misgivings are understandable. Brown (2000) suggests two post-development procedures to provide further evidence of construct validity for the scale. The first procedure takes the form of a differential-groups study, where performances on the test are compared between two

groups; one group is identified as having the construct and one group is identified without the construct. The second method consists of an intervention study, wherein a group that is considered weak on the construct is measured, exposed to the construct, and measured again. If a significant difference is found between the first and subsequent administration, then that difference can be used to support construct validity of the test.

Suggestions for Future Development

The validation of instruments is an ongoing process, and additional studies with samples that differ from the current sample are needed. Results of the study would be more tenable with data from a larger number of test takers from various institutions of differing sizes and regions of the country. Suggested revisions regarding non-content questions should be incorporated prior to future use of the B-TILED. Calculations to confirm the cut score are also needed.

Although accepted methods have been developed for establishing cut scores, results of the initial cut score calculation are preliminary. This is especially pertinent given that students had been exposed to varying levels of information literacy instruction. Judgmental and judgmental-empirical methods that rely on a panel of experts, possibly informed by item analysis and test data, are recommended for establishing a cut score (Berk, 1986). However, Berk adds that it is essential that performance data be used to confirm or refine the suggested cut score, and empirical-judgmental methods are suggested for adjusting the preliminary cut score.

One such statistical procedure might be Chester Harris' "index of efficiency," which is a formula used to calculate the one score which optimizes the pass/fail option (Harris, 1974b). Harris explains that if each individual in the top scoring group is assigned a true score equal to the mean of the upper group (and likewise for the lower group), then some generalizations can be made based on these values. The cut point, or passing score, is determined when the proportion in the upper (or lower) group is close to one-half and decreases as the proportion diverges from one-half. Kuyper and Dziuban (1984) have applied Harris' index of efficiency to passing scores on the Medical Record Administration Registration Examination and confirmed the cut score suitably sorted students into passing and failing categories.

Further investigation is also needed in the form of predictive validity. As more studies are conducted, attention should be paid to the capabilities of the instrument to predict academic or professional success. Harris (1974a) defines *instructional bias* as the distinction between testing to see if a student read the book and testing to see what reading the book did to the student. He concludes that these procedures are fundamental to determining the success of transferability of cognitive skills to behavior in authentic environments. Harris' comments draw attention to the limitations of a selected response achievement test while suggesting a look at performance on authentic tasks.

Summary

This study resulted in an instrument that is easily administered and scored that can be used to assess education students' information literacy levels. Results are significant

for reasons that range from theory-building to practical application. Understanding how instruction impacts information literacy skills levels is a necessary first step to developing a theory-connected practice of effective instructional techniques. Considerable scope exists to make use of this instrument in replicating information literacy instruction assessment across different institutional settings. It is expected that use of scales that have undergone rigorous scrutiny, such as the B-TILED, will lead to more systematic assessment of instruction and more credible reporting in the literature.

The primary goal of the study, however, was much more practical in nature. Simply put, the expectation is the test will be used to measure education students' information literacy skill levels. However, how results are analyzed, interpreted, and applied are dependent upon the reason for assessment. While individual scores can be used to identify a student's progress, cohort scores may provide more valuable data by supplying a quantitative measure of outcomes based assessment for curriculum and instruction decisions and accreditation purposes. The instrument can be used for purposes that rely on cohort scores, as well as assessing individual student mastery.

APPENDIX A:
IRB-APPROVAL AND PARTICIPANT VOLUNTEER FORM

DEVELOPMENT AND VALIDATION OF AN EDUCATION INFORMATION LITERACY INSTRUCTION TEST

Volunteer Form

You are being asked to participate in a study that is attempting to design and test an instrument that will be used to assess education students' library information literacy skills. Participation in the study will entail completing questions regarding demographic information and a draft version of the assessment instrument. It is expected this study will result in a standardized measure that can be used to assess information literacy levels and evaluate information literacy programs.

Every precaution will be made to keep participants' identities anonymous. Any identifying information, beyond demographic data, will be stripped after entering results into the database. No faculty will know if you decide to participate or decline and your decision will have no impact upon your grades. If you choose not to volunteer for this study there will be no repercussions. You may also withdraw from the study at any time.

State Limited Liability Statement

If you believe you have been injured during participation in this research project, you may file a claim against the State of Florida by filing a claim with the University of Central Florida's Insurance Coordinator, Purchasing Department, 4000 Central Florida Boulevard, Suite 360, Orlando, FL 32816, (407) 823-2661. University of Central Florida is an agency of the State of Florida and that the university's and the state's liability for personal injury or property damage is extremely limited under Florida law. Accordingly, the university's and the state's ability to compensate you for any personal injury or property damage suffered during this research project is very limited.

Information regarding your rights as a research volunteer may be obtained from:

Chris Grayson
Institutional Review Board (IRB)
University of Central Florida (UCF)
12443 Research Parkway, Suite 207
Orlando, Florida 32826-3252
Telephone: (407) 823-2901

Your signature below indicates your willingness to participate in this study. If you agree to participate, you will be expected to complete the survey to the best of your ability.

PRINT NAME:_____

SIGNATURE:_____

University of Central Florida

Office of Research

September 10, 2004

Penny Beile
University of Central Florida
Curriculum Materials Center, UCF Libraries
Orlando, FL 32816-2666

Dear Ms. Beile:

With reference to your protocol entitled, "Development and Validation of a Standards-Based Instrument for Assessing Pre-Service Teachers' Information Literacy Levels" I am enclosing for your records the approved, expedited document of the UCFIRB Form you had submitted to our office.

Please be advised that this approval is given for one year. Should there be any addendums or administrative changes to the already approved protocol, they must also be submitted to the Board. Changes should not be initiated until written IRB approval is received. Adverse events should be reported to the IRB as they occur. Further, should there be a need to extend this protocol, a renewal form must be submitted for approval at least one month prior to the anniversary date of the most recent approval and is the responsibility of the investigator (UCF).

Should you have any questions, please do not hesitate to call me at 407-823-2901.

Please accept our best wishes for the success of your endeavors.

Cordially,

Barbara Ward

Barbara Ward, CIM
IRB Coordinator

Copies: IRB File
Dr. David Boote, Educational Studies, Room 223H

12443 Research Parkway • Suite 302 • Orlando, FL 32826-3252 • 407-823-3778 • Fax 407-823-3299

APPENDIX B:
ACRL INFORMATION LITERACY COMPETENCY STANDARDS FOR HIGHER EDUCATION

ACRL Information Literacy Competency Standards for Higher Education

Standards, Performance Indicators, and Outcomes

Standard One

The information literate student determines the nature and extent of the information needed.
Performance Indicators:

1. The information literate student defines and articulates the need for information.

   Outcomes Include:

   a. Confers with instructors and participates in class discussions, peer workgroups, and electronic discussions to identify a research topic, or other information need

   b. Develops a thesis statement and formulates questions based on the information need

   c. Explores general information sources to increase familiarity with the topic

   d. Defines or modifies the information need to achieve a manageable focus

   e. Identifies key concepts and terms that describe the information need

   f. Recognizes that existing information can be combined with original thought, experimentation, and/or analysis to produce new information

2. The information literate student identifies a variety of types and formats of potential sources for information.

   Outcomes Include:

   a. Knows how information is formally and informally produced, organized, and disseminated

   b. Recognizes that knowledge can be organized into disciplines that influence the way information is accessed

   c. Identifies the value and differences of potential resources in a variety of formats (e.g., multimedia, database, website, data set, audio/visual, book)

   d. Identifies the purpose and audience of potential resources (e.g., popular vs. scholarly, current vs. historical)

e. Differentiates between primary and secondary sources, recognizing how their use and importance vary with each discipline

f. Realizes that information may need to be constructed with raw data from primary sources

3. The information literate student considers the costs and benefits of acquiring the needed information.

   Outcomes Include:

   a. Determines the availability of needed information and makes decisions on broadening the information seeking process beyond local resources (e.g., interlibrary loan; using resources at ther locations; obtaining images, videos, text, or sound)

   b. Considers the feasibility of acquiring a new language or skill (e.g., foreign or discipline-based) in order to gather needed information and to understand its context

   c. Defines a realistic overall plan and timeline to acquire the needed information

4. The information literate student reevaluates the nature and extent of the information need.

   Outcomes Include:

   a. Reviews the initial information need to clarify, revise, or refine the question

   b. Describes criteria used to make information decisions and choices

Standard Two

The information literate student accesses needed information effectively and efficiently.

Performance Indicators:

1. The information literate student selects the most appropriate investigative methods or information retrieval systems for accessing the needed information.

Outcomes Include:

a. Identifies appropriate investigative methods (e.g., laboratory experiment, simulation, fieldwork)

b. Investigates benefits and applicability of various investigative methods

c. Investigates the scope, content, and organization of information retrieval systems

d. Selects efficient and effective approaches for accessing the information needed from the investigative method or information retrieval system

2. The information literate student constructs and implements effectively-designed search strategies.

Outcomes Include:

a. Develops a research plan appropriate to the investigative method

b. Identifies keywords, synonyms and related terms for the information needed

c. Selects controlled vocabulary specific to the discipline or information retrieval source

d. Constructs a search strategy using appropriate commands for the information retrieval system selected (e.g., Boolean operators, truncation, and proximity for search engines; internal organizers such as indexes for books)

e. Implements the search strategy in various information retrieval systems using different user interfaces and search engines, with different command languages, protocols, and search parameters

f. Implements the search using investigative protocols appropriate to the discipline

3. The information literate student retrieves information online or in person using a variety of methods.

Outcomes Include:

a. Uses various search systems to retrieve information in a variety of formats

b. Uses various classification schemes and other systems (e.g., call number systems or indexes) to locate information resources within the library or to identify specific sites for physical exploration

c. Uses specialized online or in person services available at the institution to retrieve information needed (e.g., interlibrary loan/document delivery, professional associations, institutional research offices, community resources, experts and practitioners)

d. Uses surveys, letters, interviews, and other forms of inquiry to retrieve primary information

4. The information literate student refines the search strategy if necessary.

   Outcomes Include:

   a. Assesses the quantity, quality, and relevance of the search results to determine whether alternative information retrieval systems or investigative methods should be utilized

   b. Identifies gaps in the information retrieved and determines if the search strategy should be revised

   c. Repeats the search using the revised strategy as necessary

5. The information literate student extracts, records, and manages the information and its sources.

   Outcomes Include:

   a. Selects among various technologies the most appropriate one for the task of extracting the needed information (e.g., copy/paste software functions, photocopier, scanner, audio/visual equipment, or exploratory instruments)

   b. Creates a system for organizing the information

   c. Differentiates between the types of sources cited and understands the elements and correct syntax of a citation for a wide range of resources

   d. Records all pertinent citation information for future reference

   e. Uses various technologies to manage the information selected and organized

Standard Three

The information literate student evaluates information and its sources critically and incorporates selected information into his or her knowledge base and value system.

Performance Indicators:

1. The information literate student summarizes the main ideas to be extracted from the information gathered.

   Outcomes Include:

   a. Reads the text and selects main ideas

   b. Restates textual concepts in his/her own words and selects data accurately

   c. Identifies verbatim material that can be then appropriately quoted

2. The information literate student articulates and applies initial criteria for evaluating both the information and its sources.

   Outcomes Include:

   a. Examines and compares information from various sources in order to evaluate reliability, validity, accuracy, authority, timeliness, and point of view or bias

   b. Analyzes the structure and logic of supporting arguments or methods

   c. Recognizes prejudice, deception, or manipulation

   d. Recognizes the cultural, physical, or other context within which the information was created and understands the impact of context on interpreting the information

3. The information literate student synthesizes main ideas to construct new concepts.

   Outcomes Include:

   a. Recognizes interrelationships among concepts and combines them into potentially useful primary statements with supporting evidence

   b. Extends initial synthesis, when possible, at a higher level of abstraction to construct new hypotheses that may require additional information

   c. Utilizes computer and other technologies (e.g. spreadsheets, databases, multimedia, and audio or visual equipment) for studying the interaction of ideas and other phenomena

4. The information literate student compares new knowledge with prior knowledge to determine the value added, contradictions, or other unique characteristics of the information.

Outcomes Include:

a. Determines whether information satisfies the research or other information need

b. Uses consciously selected criteria to determine whether the information contradicts or verifies information used from other sources

c. Draws conclusions based upon information gathered

d. Tests theories with discipline-appropriate techniques (e.g., simulators, experiments)

e. Determines probable accuracy by questioning the source of the data, the limitations of the information gathering tools or strategies, and the reasonableness of the conclusions

f. Integrates new information with previous information or knowledge

g. Selects information that provides evidence for the topic

5. The information literate student determines whether the new knowledge has an impact on the individual's value system and takes steps to reconcile differences.

   Outcomes Include:

   a. Investigates differing viewpoints encountered in the literature

   b. Determines whether to incorporate or reject viewpoints encountered

6. The information literate student validates understanding and interpretation of the information through discourse with other individuals, subject-area experts, and/or practitioners.

   Outcomes Include:

   a. Participates in classroom and other discussions

   b. Participates in class-sponsored electronic communication forums designed to encourage discourse on the topic (e.g., email, bulletin boards, chat rooms)

   c. Seeks expert opinion through a variety of mechanisms (e.g., interviews, email, listservs)

7. The information literate student determines whether the initial query should be revised.

Outcomes Include:

a. Determines if original information need has been satisfied or if additional information is needed

b. Reviews search strategy and incorporates additional concepts as necessary

c. Reviews information retrieval sources used and expands to include others as needed

Standard Four

The information literate student, individually or as a member of a group, uses information effectively to accomplish a specific purpose.

Performance Indicators:

1. The information literate student applies new and prior information to the planning and creation of a particular product or performance.

   Outcomes Include:

   a. Organizes the content in a manner that supports the purposes and format of the product or performance (e.g. outlines, drafts, storyboards)

   b. Articulates knowledge and skills transferred from prior experiences to planning and creating the product or performance

   c. Integrates the new and prior information, including quotations and paraphrasings, in a manner that supports the purposes of the product or performance

   d. Manipulates digital text, images, and data, as needed, transferring them from their original locations and formats to a new context

2. The information literate student revises the development process for the product or performance.

   Outcomes Include:

   a. Maintains a journal or log of activities related to the information seeking, evaluating, and communicating process

   b. Reflects on past successes, failures, and alternative strategies

3. The information literate student communicates the product or performance effectively to others.

   Outcomes Include:

   a. Chooses a communication medium and format that best supports the purposes of the product or performance and the intended audience

   b. Uses a range of information technology applications in creating the product or performance

   c. Incorporates principles of design and communication

   d. Communicates clearly and with a style that supports the purposes of the intended audience

Standard Five

The information literate student understands many of the economic, legal, and social issues surrounding the use of information and accesses and uses information ethically and legally.

Performance Indicators:

1. The information literate student understands many of the ethical, legal and socio-economic issues surrounding information and information technology.

   Outcomes Include:

   a. Identifies and discusses issues related to privacy and security in both the print and electronic environments

   b. Identifies and discusses issues related to free vs. fee-based access to information

   c. Identifies and discusses issues related to censorship and freedom of speech

   d. Demonstrates an understanding of intellectual property, copyright, and fair use of copyrighted material

2. The information literate student follows laws, regulations, institutional policies, and etiquette related to the access and use of information resources.

   Outcomes Include:

a. Participates in electronic discussions following accepted practices (e.g. "Netiquette")

b. Uses approved passwords and other forms of ID for access to information resources

c. Complies with institutional policies on access to information resources

d. Preserves the integrity of information resources, equipment, systems and facilities

e. Legally obtains, stores, and disseminates text, data, images, or sounds

f. Demonstrates an understanding of what constitutes plagiarism and does not represent work attributable to others as his/her own

g. Demonstrates an understanding of institutional policies related to human subjects research

3. The information literate student acknowledges the use of information sources in communicating the product or performance.

   Outcomes Include:

   a. Selects an appropriate documentation style and uses it consistently to cite sources

   b. Posts permission granted notices, as needed, for copyrighted material

Retrieved June 4, 2004, from
http://www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.htm

APPENDIX C:
ISTE NATIONAL EDUCATIONAL TECHNOLOGY STANDARDS FOR TEACHERS (NETS*T)

ISTE Recommended Foundations

in Technology for All Teachers

The ISTE Foundation Standards reflect professional studies in education that provide fundamental concepts and skills for applying information technology in educational settings. All candidates seeking initial certification or endorsements in teacher preparation programs should have opportunities to meet the educational technology foundations standards. The following are the approved ISTE NETS for Teachers Standards.

Educational Technology Standards and Performance Indicators for All Teachers

Building on the NETS for Students, the ISTE NETS for Teachers (NETS•T), which focus on preservice teacher education, define the fundamental concepts, knowledge, skills, and attitudes for applying technology in educational settings. All candidates seeking certification or endorsements in teacher preparation should meet these educational technology standards. It is the responsibility of faculty across the university and at cooperating schools to provide opportunities for teacher candidates to meet these standards.

The six standards areas with performance indicators listed below are designed to be general enough to be customized to fit state, university, or district guidelines and yet specific enough to define the scope of the topic. Performance indicators for each standard provide specific outcomes to be measured when developing a set of assessment tools.The standards and the performance indicators also provide guidelines for teachers currently in the classroom .

I. TECHNOLOGY OPERATIONS AND CONCEPTS.

Teachers demonstrate a sound understanding of technology operations and concepts. Teachers:

A. demonstrate introductory knowledge, skills, and understanding of concepts related to technology (as described in the ISTE National Education Technology Standards for Students)

B. demonstrate continual growth in technology knowledge and skills to stay abreast of current and emerging technologies.

II. PLANNING AND DESIGNING LEARNING ENVIRONMENTS AND EXPERIENCES.

Teachers plan and design effective learning environments and experiences supported by technology. Teachers:

A. design developmentally appropriate learning opportunities that apply technology-enhanced instructional strategies to support the diverse needs of learners.

B. apply current research on teaching and learning with technology when planning learning environments and experiences.

C. identify and locate technology resources and evaluate them for accuracy and suitability.

D. plan for the management of technology resources within the context of learning activities.

E. plan strategies to manage student learning in a technology-enhanced environment.

III. TEACHING, LEARNING, AND THE CURRICULUM.

Teachers implement curriculum plans, that include methods and strategies for applying technology to maximize student learning. Teachers:

A. facilitate technology-enhanced experiences that address content standards and student technology standards.

B. use technology to support learner-centered strategies that address the diverse needs of students.

C. apply technology to develop students' higher order skills and creativity.

D. manage student learning activities in a technology-enhanced environment.

IV. ASSESSMENT AND EVALUATION.

Teachers apply technology to facilitate a variety of effective assessment and evaluation strategies. Teachers:

A. apply technology in assessing student learning of subject matter using a variety of assessment techniques.

B. use technology resources to collect and analyze data, interpret results, and communicate findings to improve instructional practice and maximize student learning.

C. apply multiple methods of evaluation to determine students' appropriate use of technology resources for learning,communication,and productivity.

V. PRODUCTIVITY AND PROFESSIONAL PRACTICE.

Teachers use technology to enhance their productivity and professional practice. Teachers:

A. use technology resources to engage in ongoing professional development and lifelong learning.

B. continually evaluate and reflect on professional practice to make informed decisions regarding the use of technology in support of student learning.

C. apply technology to increase productivity.

D. use technology to communicate and collaborate with peers, parents, and the larger community in order to nurture student learning.

VI. SOCIAL, ETHICAL, LEGAL, AND HUMAN ISSUES.

Teachers understand the social,ethical,legal,and human issues surrounding the use of technology in PK-12 schools and apply those principles in practice. Teachers:

A. model and teach legal and ethical practice related to technology use.

B. apply technology resources to enable and empower learners with diverse backgrounds, characteristics, and abilities.

C. identify and use technology resources that affirm diversity

D. promote safe and healthy use of technology resources.

E. facilitate equitable access to technology resources for all students.

Retrieved June 4, 2004, from http://cnets.iste.org/ncate/n_found.html

APPENDIX D:
MAP OF NETS*T CONTENT CLUSTERS WITH ACRL OBJECTIVES

The following content clusters illustrate the link between NETS*T content clusters and specific ACRL information literacy objectives.

Cluster 1: Identifying, Evaluating, and Selecting Finding Tools Objectives

ACRL
2.3.1.2 Identifies research sources, regardless of format, that are appropriate to a particular discipline or research need.
1.2.2.4 Describes how the publication cycle in a particular discipline or subject field affects the researcher's access to information.
1.1.3.2 Demonstrates when it is appropriate to use a general and subject-specific information source (e.g., to provide an overview, to give ideas on terminology).
1.1.4.5 Uses (describes when to use) background information sources effectively to gain an initial understanding of the topic.
2.1.3.5 Selects appropriate tools (e.g., indexes, online databases) for research on a particular topic.
2.2.6.1 Locates (identifies) major print bibliographic and reference sources appropriate to the discipline of a research topic.
2.1.3.4 Distinguishes among indexes, online databases, and collections of online databases, as well as gateways to different databases and collections.
2.2.6.2 Locates and uses (identifies) a specialized dictionary, encyclopedia, bibliography, or other common reference tool in print format for a given topic.
2.3.1.4 Uses (discriminates among) different research sources (e.g., catalogs and indexes) to find different types of information (e.g., books and periodical articles).
2.3.2.2 Explains (Indicates) the difference between the library catalog and a periodical index.
1.2.3.1 Identifies various formats in which information is available.
2.1.3 Investigates (Identifies) the scope, content, and organization of information retrieval systems
2.1.3.10 Demonstrates when it is appropriate to use a single tool (e.g., using only a periodical index when only periodical articles are required).
2.3.2.3 Describes the different scopes of coverage found in different periodical indexes.
1.1.4.6 (Identifies who to) Consults with the course instructor and librarians to develop a manageable focus for the topic.

* The objective as written reflects the original statement. Verbs in parentheses have been changed from the original to indicate measurable behaviors.

Cluster 2: Demonstrating Knowledge of General Search Strategies Objectives

ACRL
1.2.2.3 Uses* (Identifies) relevant subject- and discipline-related terminology in the information research process. 2.2.2.3 Identifies alternate terminology, including synonyms, broader or narrower words and phrases that describe a topic. 1.1.5.3 Decides (indicates) when a research topic has multiple facets or may need to be put into a broader context. 1.4.1.1 Identifies a research topic that may require revision, based on the amount of information found (or not found). 3.7.2.1 Demonstrates how searches may be limited or expanded by modifying search terminology or logic. 2.1.3.7 Identifies and uses search language and protocols (e.g., Boolean, adjacency) appropriate to the retrieval system. 2.2.4.2 Demonstrates an understanding of the concept of Boolean logic and constructs a search statement using Boolean operators. 2.2.4.3 Demonstrates an understanding of the concept of proximity searching and constructs a search statement using proximity operators. 2.2.4.6 Demonstrates an understanding of the concept of keyword searching and uses it appropriately and effectively. 2.2.4.7 Demonstrates an understanding of the concept of truncation and uses it appropriately and effectively. 2.2.4.1 Demonstrates when it is appropriate to search a particular field (e.g., title, author, subject). 2.2.5.2 Demonstrates an awareness of the fact that there may be separate interfaces for basic and advanced searching in retrieval systems. 2.2.5.3 Narrows or broadens questions and search terms to retrieve the appropriate quantity of information, using search techniques such as Boolean logic, limiting, and field searching. 2.3.1.5 Describes search functionality common to most databases regardless of differences in the search interface (e.g., Boolean logic capability, field structure, keyword searching, Relevancy ranking). 3.7.3.1 Examines footnotes and bibliographies from retrieved items to locate additional sources. (?) 2.2.3.1 Uses (distinguishes among) background sources (e.g., encyclopedias, handbooks, dictionaries, thesauri, textbooks) to identify discipline-specific terminology that describes a given topic. 2.2.3.2 Explains what controlled vocabulary is and why it is used. 2.2.3.3 Identifies search terms likely to be useful for a research topic in relevant controlled vocabulary lists. 2.2.3.4 Identifies when and where controlled vocabulary is used in a bibliographic record, and then successfully searches for additional information using that vocabulary.

* The objective as written reflects the original statement. Verbs in parentheses have been changed from the original to indicate measurable behaviors.

Cluster 3: Evaluating and Selecting Sources Objectives

ACRL
2.3.1.3 Recognizes (distinguishes) the format of an information source (e.g., book, chapter in a book, periodical article) from its citation. 2.3.2.4 Distinguishes among citations to identify various types of materials (e.g., books, periodical articles, essays in anthologies). 2.5.3.1 Identifies different types of information sources cited in a research tool. 5.3.1.2 Identifies citation elements for information sources in different formats (e.g., book, article, television program, Web page, interview). 2.4.1.3 Assesses the relevance of information found by examining elements of the citation such as title, abstract, subject headings, source, and date of publication. 2.4.1.2 Evaluates (explains) the quality of the information retrieved using criteria such as authorship, point of view/bias, date written, citations, etc. 3.2.1.2 (Indicates when to) Investigate an author's qualifications and reputation through reviews or biographical sources. 3.2.1.4 (Indicates when to) Investigate qualifications and reputation of the publisher or issuing agency by consulting other information resources. 3.2.3.1 Demonstrates an understanding that information in any format reflects an author's, sponsor's, and/or publisher's point of view.

* The objective as written reflects the original statement. Verbs in parentheses have been changed from the original to indicate measurable behaviors.

Goal 4: Demonstrating Knowledge of Legal and Ethical Practices Objectives

ACRL
5.1.4 Demonstrates an understanding of intellectual property, copyright, and fair use of copyrighted material 5.2.6 Demonstrates an understanding of what constitutes plagiarism and does not represent work attributable to others as his/her own

APPENDIX E:
TEST ITEMS WITH ALIGNMENT TO NETS*T CONTENT CLUSTERS AND ACRL INFORMATION LITERACY STANDARDS OBJECTIVES

item #	item	NETS*T	ACRL
1	You have been assigned to write a paper on the whole language movement, a topic with which you are unfamiliar. Which of the following is the best source to find a brief history and summary? (Check only one answer.) A. a book titled Perspectives on Whole Language Learning: A Case Study b. a dissertation titled Whole language and learning disabilities: Case study of a student teacher's beliefs development c. a recent newspaper article titled "Whole Language in the Classroom: How Effective?" d. Encyclopedia Britannica e. Encyclopedia of Education	1	1.1.3.2
2	While observing a class, you hear one teacher mention to another that they are designing their environmental science unit using a "constructivist" approach. You are not certain what this means, so decide to: (Check only one answer.) a. search for background information on the topic b. search for editorials on the topic c. search for literature reviews on the topic d. search for research reports on the topic	1	1.1.4.5
3	What are the best things to do when you need help focusing your research topic? (Check all that apply.) A. Ask a person in the library who is shelving books b. Ask a person in the library who is staffing the circulation desk c. Ask a person in the library who is staffing the reference desk d. Ask your course instructor e. Seek assistance through your library's Web-based assistance service	1	1.1.4.6
4	You have been assigned a comprehensive research paper on the impact of Title IX on high school athletics, which essentially served to restructure school athletics by mandating financial and technical support for girls sports programs. Which of the following strategies would you use to locate information? (Check only one answer.) a. Searching a number of general and discipline sources is needed. b. Searching education sources alone should be sufficient. c. Searching general sources alone should be sufficient. d. Searching government documents alone should be sufficient.	2	1.1.5.3
5	Select the set of search terms that best represents the main concepts in the following sentence. What are the health risks associated with the use of anabolic steroids and women athletes? (Check only one answer.) A. drugs, athletes, performance b. girls, females, women c. performance enhancing drugs, women d. Women athletes, health risks, steroids e. Women, females, girls	2	1.2.2.3

6	Generally, research in education is initially communicated through: (Check only one answer.) a. books published by university presses b. education encyclopedia entries c. Internet Web-logs d. newsletters of professional organizations e. professional conferences and journal articles	1	1.2.2.4
7	Most recent ERIC documents are available: (Check all that apply.) a. as proceedings in conference reports b. in books c. In education journals d. in electronic format e. In microfiche	1	1.2.3.1
8	When searching ERIC for information on inclusion, what is the best strategy when a search yields over a thousand (>1000) results? (Check only one answer.) a. add another term to the search b. change your topic to something else c. look at all of the results, to not miss something good	2	1.4.1.1
9	By design, items included in a research database are limited to which of the following criteria? (Check all that apply.) A. Date range b. not found on the Internet c. Not copyrighted d. Owned by your library e. Subject matter	1	2.1.3
10	Which of the following statements are generally true about a Web search engine (for example, Google, Altavista, etc.)? (Check all that apply.) A. Searches free resources b. Searches for peer-reviewed materials c. Searches most research databases d. Searches material not found in print e. uses subject headings	1	2.1.3
11	You have been assigned a short class paper on teaching ESL students in the classroom. Your professor indicated three recent scholarly sources would be sufficient. Which strategy is best to locate sources? (Check only one answer.) A. search library catalogs for encyclopedias b. search multiple databases for newspapers c. search periodical databases for journal articles d. Search multiple databases for books, encyclopedias, and journal articles e. Search multiple databases for journal articles and newspapers	1	2.1.3.10
12	Which of the following is a gateway to different research databases? (Check only one answer.) a. Academic Search Premier b. EBSCOhost c. ERIC d. Professional Development Collection e. PsycINFO	1	2.1.3.4
13	You need to find journal articles on the topic of gender differences in math and science achievement to complete a paper. Which of these would be a good way to start? (Check only one answer.) a. Page through recent journals b. use a research database c. use a Web search engine d. use the library catalog	1	2.1.3.5

14	Research databases vary in their search protocols. For example, one database may use an asterik (*) as a truncation symbol while another database uses a question mark (?). How do you identify search protocols appropriate to the retrieval system? (Check all that apply.) a. look at the database's search help screen b. if offered, work through the database's tutorial on searching c. type in different symbols until good results are received	2	2.1.3.7
15	Select the set of terms that best represents a partial list of synonyms and related terms for the concept "college students." (Check only one answer.) a. Colleges, universities, freshmen, community colleges… b. undergraduates, graduate students, freshmen, sophomores... c. university, adult learners, students, educational attendees... d. Young people, kids, youth, young adults…	2	2.2.2.3
16	You need to find background information on the Montessori movement. A good place to look would be: (Check only one answer.) a. Digest of Education Statistics b. Education Almanac c. Encyclopedia of Education d. Learning and Instruction textbook e. Thesaurus of ERIC Descriptors	2	2.2.3.1
17	The definition of the term "controlled vocabulary" is: (Check only one answer.) a. A part of a record (such as author, title, abstract). b. A subject heading assigned to an article or document. C. A summary of an article or document. d. The citation and abstract fields	2	2.2.3.2
18	Typing the term school vouchers into a research database thesaurus returned the following suggestions. Which of the results is likely to be most useful for your search? (Check only one answer.) a. educational equality b. private schools c. school choice d. state aid to education e. tuition fees	2	2.2.3.3
19	How can you find good subject headings for articles on your topic in a research database? (Check all that apply.) A. Try a keyword search to locate a good article and look at its subject headings b. Use the index in a book c. Use the subject categories from Yahoo! d. Use the thesaurus for the research database	2	2.2.3.4
20	If you wanted to find information witten about educator John Dewey, which search would you do? (Check only one answer.) A. author: Dewey b. Subject: Dewey c. Title: Dewey	2	2.2.4.1

21	You have found the following reference: Shayer, M. (2003). Not just Piaget, not just Vygotsky. Learning and Instruction, 13(5), 465-485. Which of the following would you type into the library's catalog to locate the actual article? (Check only one answer.) A. Learning and Instruction as a title search B. Not just Piaget, not just Vygotsky as a title search C. Piaget and Vygotsky as a keyword search d. Piaget and Vygotsky as an author search e. Shayer as an author search	2	2.2.4.1
22	If you are researching a paper on treatment for hyperactivity, which of the following keyword searches is the most appropriate? (Check only one answer.) A. Hyperactivity treatment b. Treatment and hyperactivity c. Treatment for hyperactivity d. treatment of hyperactivity d. Treatment or hyperactivity	2	2.2.4.2
23	Which of the following search statements would retrieve more records? A. Behavior disorders adj hyperactivity b. Behavior disorders and hyperactivity c. Behavior disorders or hyperactivity d. Behavior disorders not hyperactivity e. "behavior disorders and hyperactiviity"	2	2.2.4.2
24	You are searching for articles on high school persistence and you want the words high school to occur beside each other. Which of the following keyword searches would be most appropriate? (Check only one answer.) A. High and school and persistence b. Persistence and high adj school c. Persistence and high school d. Persistence or high adj school	2	2.2.4.3
25	You want to locate information on student plagiarism. If you type in the term as a keyword search, what part of the record is being searched? (Check all that apply.) a. Abstract or contents field b. Author field c. subject headings field d. Title field	2	2.2.4.6
26	If you end a search term with a special symbol like this: child* or child!, you would retrieve articles that contain which of the following words? (Check all that apply.) A. Child b. Childbirth c. Children c. Students e. Youth	2	2.2.4.7
27	Truncation allows you to: (Check only one answer.) a. Combine different concepts b. Combine similar concepts c. Search for variant word endings d. Search for words as phrases	2	2.2.4.7
28	Most library databases have basic and advanced searching interfaces. Which of the following can you only do in advanced searching? (Check all that apply.) a. add search connectors between terms b. enter multiple search terms c. search for a single keyword d. search multiple terms by field	2	2.2.5.2

29	You are interested in whether service learning experiences in high school impact job choice, so you type service learning as a keyword search into the ERIC database and retrieve over 2,000 items. How might you narrow your search results? (Check all that apply.) a. add the term high school b. add the term job choice c. find a descriptor to search by looking at subject headings in the thesaurus d. find a descriptor to search by looking at the subject headings in a few relevant articles e. limit to items published in the last few years	2	2.2.5.3
30	You have been asked to make a presentation to your class on the educational system in Botswana. Which sources would offer the most appropriate information? (Check all that apply.) a. Education and Sociology: An Encyclopedia b. Encyclopedia of World History c. Guide to Higher Education in Africa d. World Education Encyclopedia e. World List of Universities	1	2.2.6.1
31	The best source to find the meaning of the term "bilingual education" is: (Check only one answer.) a. Dictionary of Education b. Education Yearbook C. ERIC Thesaurus D. Webster's Dictionary	1	2.2.6.2
32	Which source would provide the best overview for Piaget's child development theories as they relate to education? (Check only one answer.) a. Child Development Abstracts b. Education Full Text c. Encyclopedia Americana d. Encyclopedia of Education e. ERIC	1	2.2.6.2
33	A bibliography in a book is useful because: (Check only one answer.) a. It gives you a short summary of the author's life b. It lets you know what other people think of the work c. It lists contents of chapters in the book d. it lists the information sources used by the author e. It tells you on what pages of the book you will find the subject of interest	1	2.3.1.2
34	Where is the best place to find information about ASCD's (a professional association) activities and membership? (Check only one answer.) A. Books b. Encyclopedia c. Internet d. Journal article e. Magazine article	1	2.3.1.2
35	Based on the following citation, what would you type into the library catalog to see if it's available at your institution? (Check only one answer.) Casellon, D. C. (2000). Lunatics are running the asylum!: A study in higher education administration. Journal of Higher Education, 17(1), 171-180. a. Author search on: Casellon, D. C. b. Journal title search on: Journal of Higher Education c. Journal title search on: Lunatics are running the asylum d. Subject search on: Higher Education Administration	3	2.3.1.3

36	Given this citation: Massaro, D. (1991). Broadening the domain of the fuzzy logical model of perception. In H. L. Pick, Jr., P. van den Broek, & D. C. Knill (Eds.), Cognition: Conceptual and methodological issues (pp. 51-84). Washington, DC: American Psychological Association, "Broadening the domain of the fuzzy logical model of perception" is the: (Check only one answer.) a. Name of a book b. name of a chapter in a book c. name of a journal d. name of an article	3	2.3.1.3
37	Which database would you search to find books held by the library? (Check only one answer.) A. EbscoHOST b. Education Full Text c. ERIC d. Library catalog e. Professional Development Collection	1	2.3.1.4
38	Why would you use a periodical index or database? (Check only one answer.) A. to check your email b. To search the Web c. To find citations or articles d. To see if the library owns a journal e. To see if the library owns a magazine	1	2.3.1.4
39	In most research databases, an advantage to using a key word search is that key word searches: (Check only one answer.) a. Are especially useful for topics with an established body of literature b. Are more discriminating and yield more appropriate citations c. Search all parts of the record (author, title, abstract, etc) and yield more citations d. use Library of Congress Subject Headings, so choosing terms is easy.	2	2.3.1.5
40	What is the term for an online resource that shows what materials a library owns? (Check only one answer.) a. Bibliography b. Keyword c. Library catalog d. Research database e. Subject heading	1	2.3.2.2
41	ERIC is the most appropriate database to search to locate articles and documents concerning: (Check only one answer.) a. education statistics b. historical subjects c. research conducted by the US Department of Education d. c. research from 1965 to the present	1	2.3.2.3
42	The following citation is a/an: (Check only one answer.) a. Book b. Chapter in a book c. Journal article d. Other Elkind, D. (1978). The child's reality: Three developmental themes. New Jersey: Lawrence Erlbaum Associates.	3	2.3.2.4
43	The following citation is a/an: (Check only one answer.) a. Book b. Chapter in a book c. Journal article d. Other Massaro, D. (1991). Broadening the domain of the fuzzy logical model of perception. In H. L. Pick, Jr., P. van den Broek, & D. C. Knill (Eds.), Cognition: Conceptual and methodological issues (pp. 51-84). Washington, DC: American Psychological Association.	3	2.3.2.4

44	The following citation is a/an: (Check only one answer.) a. Book b. Chapter in a book c. Journal article d. Other Teri, L. (1982). Depression in adolescence: Its relationship to assertion and various aspects of self-image. Clincial Child Psychology, 11(2), 101-106.	3	2.3.2.4
45	When evaluating articles, which characteristic best indicates scholarly research? (Check only one answer.) A. Available in an academic library b. Indexed by ERIC c. Indexed by Google d. Reviewed by experts before publication e. Written by university faculty	3	2.4.1.2
46	When researching a controversial topic on the Internet, such as prayer in public schools, can you determine bias of the information before reading it? (Check only one answer.) A. No, one needs to read the information to find bias b. yes, if the information is published on the Web it should be unbiased c. yes, if the Web site is reporting research it should be unbiased d. Yes, reputation of the Web site's address, or URL, should give an indication of bias e. yes, the title of the information should indicate bias	3	2.4.1.2
47	What is the most efficient way to determine whether or not an article that you retrieved from a full-text database search is relevant to your topic? (Check only one answer.) A. Read the abstract b. Read the discussion section c. Read the introduction d. Read the methods section e. Read the results section	3	2.4.1.3
48	Who is the intended audience for this article? Title: Helping kids think and work on their own. Pages: 27-30 Abstract: Presents several techniques that can boost students' confidence levels, increase independent work habits, and help maximize communication skills on school test day and beyond. Creation of a supportive classroom environment; Promotion of student independence; Cultivation of clear thinking. (Check only one answer.) a. parents b. teachers c. general public d. scholars	3	2.4.1.3
49	You have been asked to make a presentation to your class on assistive technologies for visually challenged students. From which source would you be most likely to find appropriate information? A. Elementary School Journal b. International Journal for the Education of the Blind c. Journal of Educational Gerontology d. Journal of Military Service Learning e. Reader's Digest Magazine	3	2.4.1.3

50	The following citation was retrieved from an ERIC database search. What kind of source is it? (Check only one answer.) Title: Preservice Elementary Teachers' Self-Efficacy Beliefs Author(s): Cakiroglu, Jale; Boone, William J. Publication Year: 2001 Abstract: The purpose of this study was to examine pre-service elementary teachers' self-efficacy beliefs in teaching science. Notes: Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001). Number of Pages: 24 ERIC Number: ED453084 a. A book b. A book chapter c. A book review d. a conference paper e. A journal article	3	2.5.3.1
51	You are discussing the controversial book Cultural Literacy in your class and want to find out more about the qualifications of the author, E. D. Hirsch. A good way to go about this is: (Choose all that apply.) a. to locate information on the author from biographical sources b. to look at author information in the book c. to look at author information on the Web d. to search for reviews of the book	3	3.2.1.2
52	While developing a lesson plan on the U.S. legislative system, you find this story on the Internet: Congress Launches National Congress-Awareness Week WASHINGTON, DC—Hoping to counter ignorance of the national legislative body among U.S. citizens, congressional leaders named the first week in August National Congress Awareness Week. "This special week is designed to call attention to America's very important federal lawmaking body," Speaker of the House Dennis Hastert said. The festivities will kick off with a 10-mile Walk for Congress Awareness, when blue ribbons will be handed out in honor of those who served in the first 107 congresses. The item is from the Onion, which claims to be America's Finest News Source. Given this, the following action is in order: (Check only one answer.) a. Although the story was found on the Web, you can use it as it's obviously from a reputable news service. b. You decide to investigate the reputation of the publisher by looking at their Web site. c. You decide to investigate the reputation of the publisher by looking at other Web sites.	3	3.2.1.4

53	In completing a research paper, you do a Google search to get some general information. The first page of results shows a number of different links. Which of the following may reflect the author's, publisher's, or sponsor's point of view? (Check all that apply.) a. First page is entitled: This paper examines Direct Instruction--one branch of the "instructivist" approach in education, URL http://people.uncw.edu/kozloffm/diarticle.pdf b. What the Data Really Show: Direct Instruction Really Works! The dirty little secret from the biggest education study ever, URL:http://www.jefflindsay.com/EducData.shtml c. The Madeline Hunter Direct Instruction Model, URL: http://www.humboldt.edu/~tha1/hunter-eei.html d. Effective literacy instruction for adolescents, commissioned by the National Reading Conference,URL: http://www.nrconline.org/publications/alverwhite2.pdf	3	3.2.3.1
54	You have been assigned a paper to investigate the benefits of direct instruction as it relates to student learning. A keyword search in ERIC on direct instruction has returned 1400 items. To narrow your search, which of the following steps would you next perform? (Check only one answer.) A. Add student learning as a keyword search b. Add benefits as a subject/descriptor search c. Enter a name as an Author search d. enter a title as a Title search e. Limit search results to English	2	3.7.2.1
55	When one finds an excellent article, what is the most direct method for identifying related sources? (Check only one answer.) a. ask the course instructor b. ask a reference librarian c. look at the bibliography from the article d. search ERIC under the author's name e. search the library catalog	2	3.7.3.1
56	When is it ethical to use the ideas of another person in a research paper? (Check only one answer.) A. It is never ethical to use someone else's ideas b. Only if you do not use their exact words c. only when you give them credit d. Only when you receive their permission e. Only when you use their exact words.	4	5.1.4
57	All of the following are good strategies for avoiding plagiarism, EXCEPT: (Check only one answer.) a. Document your source by citing the information you use b. Paraphrase the idea rather than directly copying it c. Take accurate notes as to where you found specific ideas d. Use quotation marks around exact quotations	4	5.2.6

58	Using this result from an Internet search engine, who is the "owner" of this Web site? State policies on planning, funding, and standards. Does the state have technology requirements for students? http://www.edweek.org/reports/tc98/states/fl.htm (Check only one answer.) a. business entity b. college or university c. national government agency d. other organization e. state government agency	3	5.3.1.2
59	Based on the following paragraph, which sentence should be cited? (1)Technology use in the schools is often characterized as a potentially dehumanizing force. (2)Perhaps the fear that the virtual world may lead to passivity and isolation, at the expense of literal social interaction, is valid. (3)Certainly, educators must ask which uses of technology result in increased learning and a better quality of life. (4)To address these issues, one model has been proposed that suggests students work in groups with the computer peripheral to the group and the teacher serves as facilitator. CHECK ONLY ONE ANSWER. a. 1 b. 2 c. 3 d. 4	4	5.2.6
60	You are creating a Web page for a student education organization. Browsing the Internet, you find a neat photo from the US Department of Education, which is a government agency. If you decide to use the graphic on your Web page, which of the following copyright choices is the proper action? CHECK ONLY ONE ANSWER. a. Permission is not needed as the photo is from a government agency. b. Permission is not needed as the photo was found on the Internet. c. Permission is not needed as you are only using it for a Web page. d. Permission to use the photo must be acquired before using it.	4	5.1.4
61	You have an assignment that requires you to use course management software to practice setting up a class grade book. Your school has purchased the software and loaded it in the computer lab, but you have a difficult time getting to the lab due to work conflicts. A friend loans you the software and you load it on your computer. Is this legal? CHECK ONLY ONE ANSWER. a. No, because this action constitutes a violation of copyright. b. Yes, because it is already freely available in the lab. c. Yes, because it is education software and therefore able to be shared. d. Yes, because your friend owns it and can share as he wants.	4	5.2.5

62	Browsing a weekly news magazine, you come across an article that discusses the future of space exploration. As you are teaching this topic you decide to make 25 copies of the article and share it with your class. Which of the following concepts makes it legally permissible to reproduce portions of works for educational purposes without permission? CHECK ONLY ONE ANSWER. a. Copyright b. Fair use c. Freedom of information d. Intellectual freedom e. Intellectual property	4	5.1.4

APPENDIX F:
INITIAL 62 ITEMS USED FOR SMALL GROUP ADMINISTRATION

LIBRARY AND INFORMATION SKILLS QUIZ –
With Answers Underlined

1. Your professor has assigned a paper on the whole language movement. You are not familiar with the topic, so you decide to read a brief history and summary about it. Which of the following sources would be best?
CHECK ONLY ONE ANSWER.
☐ a book on the topic, such as *Perspectives on Whole Language Learning: A Case Study*
☐ a dissertation on the topic, such as *Whole language and learning disabilities: Case study of a student teacher's beliefs development*
☐ a general encyclopedia, such as *Encyclopedia Britannica*
☐ a newspaper article on the topic, such as "Whole language in the classroom: How effective?"
☐ an education encyclopedia, such as *Encyclopedia of Education*

2. While observing a class, you overhear one teacher mention to another that they are designing their environmental science unit using a "constructivist" approach. You are not certain what this means, so decide to search for::
CHECK ONLY ONE ANSWER.
☐ background information on the topic
☐ editorials on the topic
☐ literature reviews on the topic
☐ research articles on the topic

3. Who may be the most qualified person(s) to assist you when you need help narrowing your research topic?
CHECK ONLY ONE ANSWER.
☐ a fellow student in your class
☐ a person in the library who is shelving books
☐ a person in the library who is staffing the circulation desk
☐ a person in the library who is staffing the reference desk

4. You have been assigned a comprehensive (20 page) research paper on the impact of Title IX on high school sports programs. (Title IX legislation sought to ensure gender equity for sports programs.) Which of the following strategies is best to locate information?
CHECK ONLY ONE ANSWER.
☐ search for both general academic and government documents sources
☐ search for education sources only
☐ search for general academic, education, and government documents sources
☐ search for government documents sources only

5. Select the set of search terms that best represent the main concepts in the following: What are the health risks associated with the use of drug therapy for hyperactive students?
CHECK ONLY ONE ANSWER.
☐ drug therapy, health risks
☐ drugs, hyperactivity, therapy
☐ drugs, students, health risks
☐ hyperactivity, health risks, drug therapy
☐ students, hyperactivity, attention deficit disorder

6. Research studies in education are generally first communicated through:
CHECK ONLY ONE ANSWER.
☐ books published by university presses
☐ education encyclopedia entries
☐ Internet Web-logs
☐ newsletters of education associations
☐ professional conferences and journal articles

7. Most ERIC documents published since 1996 are available:
CHECK ONLY ONE ANSWER.
☐ as articles in education journals
☐ as chapters in books
☐ as electronic documents
☐ as proceedings in conference reports

8. When searching a research or periodical database, such as ERIC, for information on inclusion, which of the following is the best strategy when a search yields over a thousand (>1000) results?
CHECK ONLY ONE ANSWER.
☐ add another term and re-do the search
☐ change your topic
☐ look at all of the results

9. Research or periodical databases are designed to include items based on which of the following criteria?
CHECK ONLY ONE ANSWER.
☐ not copyrighted
☐ not found on the Internet
☐ owned by your library
☐ relevant subject matter

10. Which of the following statements is generally true about a Web search engine (for example, Google, AltaVista, etc.)?
CHECK ONLY ONE ANSWER
☐ does not search most research databases
☐ searches most research databases
☐ searches peer-reviewed materials
☐ uses controlled vocabulary to search

11. You have been assigned to write a short class paper on effective instruction techniques for teaching English as a Second Language (ESL) students. Your professor indicated three recent scholarly sources would be sufficient. Which strategy is best to locate items?
CHECK ONLY ONE ANSWER.
☐ search an education database for journal articles
☐ search several databases for journal articles and newspapers
☐ search a newspaper database for newspaper articles
☐ search the library catalog for encyclopedias

12. Which of the following choices serves as a gateway by allowing access to different periodical or research databases?
CHECK ONLY ONE ANSWER.
☐ Academic Search Premier
☐ EBSCOhost
☐ ERIC
☐ Professional Development Collection
☐ PsycINFO

13. You need to find journal articles on the topic "gender differences in math and science achievement" to write a paper. Which of the following choices is the most efficient way to start?
CHECK ONLY ONE ANSWER.
☐ page through recent journals
☐ use a research database
☐ use a Web search engine
☐ use the library catalog

14. Research databases vary in their search protocols. For example, one database may use an asterisk (*) as a truncation symbol while another database uses a question mark (?). What is the most efficient way to identify search protocols appropriate to the retrieval system?
CHECK ONLY ONE ANSWER
☐ look at the database search help screen
☐ type in different symbols until good results are received
☐ work through the database tutorial on searching

15. Select the set that best represents synonyms and related terms for the concept "college students."
CHECK ONLY ONE ANSWER.
☐ colleges, universities, community colleges…
☐ graduate students, freshmen, sophomores...
☐ university, adult learners, educational attendees...
☐ young people, youth, young adults…

16. You need to find background information on the teaching for multiple intelligences movement. Which of the following is the best source to look?
CHECK ONLY ONE ANSWER.
☐ *Digest of Education Statistics*
☐ *Education Almanac*
☐ *Encyclopedia of Education*
☐ *Learning and Instruction* textbook
☐ *Thesaurus of ERIC Descriptors*

17. The definition of the term "subject heading" is:
CHECK ONLY ONE ANSWER.
☐ a descriptor assigned to an article or document
☐ a part of a record (such as author, title, abstract)
☐ a summary of an article or document
☐ the citation and abstract fields

18. Typing the term “school vouchers” into a research or periodical database thesaurus returned the following suggested descriptors. Which of the results is likely to be most useful for your search?
CHECK ONLY ONE ANSWER.
☐ educational equality
☐ private schools
☐ school choice
☐ state aid to education
☐ tuition fees

19. How can you find good subject headings for articles on your topic in a research database?
CHECK ALL THAT APPLY.
☐ look at the subject headings in a relevant article
☐ use the index in a book
☐ use the subject categories from Yahoo!
☐ use the thesaurus for the research database

20. If you want to locate information written about educator John Dewey, which search would return the most relevant results?
CHECK ONLY ONE ANSWER.
☐ author search: Dewey
☐ keyword search: Dewey
☐ subject search: Dewey
☐ title search: Dewey

21. Your professor suggested you read a particular article and gave you the following citation: Shayer, M. (2003). Not just Piaget, not just Vygotsky. *Learning and Instruction, 13*(5), 465-485.
Which of the following would you type into the library's catalog to locate the actual article?
CHECK ONLY ONE ANSWER.
☐ author search: Piaget and Vygotsky
☐ author search: Shayer
☐ keyword search: Piaget and Vygotsky
☐ title search: Learning and Instruction
☐ title search: Not just Piaget, not just Vygotsky

22. You are researching a paper on treatment for hyperactivity, and have identified the terms "treatment" and "hyperactivity" as keywords. Which of the following Boolean connectors is the most appropriate to use between the terms?
CHECK ONLY ONE ANSWER.
☐ and
☐ for
☐ of
☐ or

23. Which of the following search statements would retrieve the most records?
CHECK ONLY ONE ANSWER.
☐ "behavior disorders and hyperactivity"
☐ behavior disorders and hyperactivity
☐ behavior disorders or hyperactivity
☐ behavior disorders not hyperactivity

24. You are searching for articles on high school persistence in a research or periodical database and you want the words "high school" to occur beside each other. Which of the following searches would be most appropriate?
CHECK ONLY ONE ANSWER.
☐ persistence adj high adj school
☐ persistence and high adj school
☐ persistence and high school
☐ persistence or high adj school

25. You want to locate information on student plagiarism. If you type in the term "plagiarism" as a keyword search, what part of the record is being searched?
CHECK ALL THAT APPLY.
☐ abstract or contents field
☐ author field
☐ subject headings field
☐ title field

26. You are using a research database that uses an asterisk (*) as its truncation symbol. When you type in "read*" you would retrieve records that contained which of the following words?
CHECK ALL THAT APPLY.
☐ examine
☐ peruse
☐ reader
☐ reading
☐ readmit

27. Truncation allows you to:
CHECK ONLY ONE ANSWER.
☐ combine different concepts
☐ combine similar concepts
☐ search for variant word endings
☐ search for words as phrases

28. Most research and periodical databases have basic and advanced searching interfaces. Which of the following can you do ONLY in advanced searching?
CHECK ONLY ONE ANSWER.
☐ add Boolean or search connectors between terms
☐ enter multiple search terms
☐ search for a single keyword
☐ search multiple terms by field

29. You are interested in whether service learning experiences in high school influence job choice, so you type "service learning" as a keyword search into the ERIC database and retrieve over 2,000 items. How might you narrow your search results?
CHECK ALL THAT APPLY.
☐ add the term "high school"
☐ add the term "job choice"
☐ look at a few relevant articles to find descriptors
☐ look at the thesaurus to find descriptors

30. You have been asked to make a presentation to your class on the PreK-12 educational system in South Africa. Which of the following sources would offer the most appropriate information?
CHECK ONLY ONE ANSWER.
☐ *Education and Sociology: An Encyclopedia*
☐ *Encyclopedia of World History*
☐ *Guide to Higher Education in Africa*
☐ *World Education Encyclopedia*
☐ *World List of Universities*

31. Of the following, which is the best source to find the meaning of the term "bilingual education?"
CHECK ONLY ONE ANSWER.
☐ *Dictionary of Education*
☐ *Education Yearbook*
☐ *Thesaurus of ERIC Descriptors*
☐ *Webster's Dictionary*

32. Which source would provide the best overview of Piaget's child development theories as they relate to education?
CHECK ONLY ONE ANSWER.
☐ *Child Development Abstracts*
☐ *Education Full Text*
☐ *Encyclopedia Americana*
☐ *ERIC*
☐ *Handbook of Educational Psychology*

33. A bibliography in a book is useful because it:
CHECK ONLY ONE ANSWER.
☐ gives you a short summary of the author's life
☐ lets you know what other people think of the work
☐ lists contents of chapters in the book
☐ lists the information sources used by the author
☐ tells you on what pages of the book you will find the subject of interest

34. Of the following choices, which is the best source to locate membership information on a professional organization, such as the Association for Supervision and Curriculum Development (ASCD)?
CHECK ONLY ONE ANSWER.
☐ books
☐ encyclopedia entries
☐ Internet
☐ journal articles
☐ magazine articles

35. Based on the following citation, what would you type into the library catalog to see if the item is available at your institution?
Casellon, D. C. (2000). Lunatics are running the asylum!: A study in higher education administration. *Journal of Higher Education, 17*(1), 171-180.
CHECK ONLY ONE ANSWER.
☐ author search: Casellon, D. C.
☐ journal title search: Journal of Higher Education
☐ journal title search: Lunatics are running the asylum
☐ subject search: higher education administration

36. In the following citation, what is "Knowing, teaching, and supervising?"
Hunter, M. (1984). Knowing, teaching, and supervising. In P. L. Hosford (Ed.), *Using what we know about teaching* (pp. 180-187). Alexandria, VA: Association for Supervision and Curriculum Development.
CHECK ONLY ONE ANSWER.
☐ a book chapter
☐ a book title
☐ a journal title
☐ an article title

37. Which online resource would you search to find books owned by your institution?
CHECK ONLY ONE ANSWER.
☐ EbscoHOST
☐ Education Full Text
☐ ERIC
☐ Library catalog
☐ Professional Development Collection

38. What is the primary reason for using a research or periodical database?
CHECK ONLY ONE ANSWER.
☐ to find citations or articles
☐ to search the Web
☐ to see if the library owns a book
☐ to see if the library owns a journal

39. In most research databases, an advantage to using a keyword search is that keyword searches:
CHECK ONLY ONE ANSWER.
☐ are especially useful for topics with an established body of literature
☐ are more discriminating and yield more appropriate citations
☐ search all parts of the record and yield more results
☐ use Library of Congress subject headings, so choosing terms is easy

40. What is the term for an online resource that shows what materials are owned by your library?
CHECK ONLY ONE ANSWER.
☐ bibliography
☐ database thesaurus
☐ library catalog
☐ periodical database
☐ research database

41. ERIC is the most appropriate database to search to locate:
CHECK ONLY ONE ANSWER.
☐ education article citations and documents
☐ education publications from 1877 to current
☐ full-text education articles
☐ US Department of Education publications
☐ world-wide education statistics

42. The following citation is for:
Elkind, D. (1978). *The child's reality: Three developmental themes*. Trenton, New Jersey: Lawrence Erlbaum Associates.
CHECK ONLY ONE ANSWER.
☐ a book
☐ a chapter in a book
☐ a journal article
☐ other

43. The following citation is for:
Massaro, D. (1991). Broadening the domain of the fuzzy logical model of perception. In H. L. Pick, Jr., P. van den Broek, & D. C. Knill (Eds.), *Cognition: Conceptual and methodological issues* (pp. 51-84). Washington, DC: American Psychological Association.
CHECK ONLY ONE ANSWER.
☐ a book
☐ a chapter in a book
☐ a journal article
☐ other

44. The following citation is for:
Teri, L. (1982). Depression in adolescence: Its relationship to assertion and various aspects of self-image. *Clinical Child Psychology, 11*(2), 101-106.
CHECK ONLY ONE ANSWER.
☐ a book
☐ a chapter in a book
☐ a journal article
☐ other

45. Which of the following characteristics best indicates scholarly research?
CHECK ONLY ONE ANSWER.
☐ available in an academic library
☐ indexed by ERIC
☐ indexed by Google
☐ reviewed by experts for publication
☐ written by university faculty

46. When researching a controversial topic on the Internet, such as prayer in public schools, can a person generally determine bias of the information before reading it?
CHECK ONLY ONE ANSWER.
☐ no, a person needs to read the information to find bias
☐ yes, if the information is published on the Web it should be unbiased
☐ yes, if the Web site is reporting research it should be unbiased
☐ yes, the title of the Web site should indicate bias

47. Assuming a research article that you retrieved from a full-text database has the following sections, which of the following would be the most efficient way to determine if it is relevant to your topic?
CHECK ONLY ONE ANSWER.
☐ read the abstract
☐ read the discussion section
☐ read the introduction
☐ read the methods section
☐ read the results section

48. Who is the intended audience for the following article?
Journal title: Teaching Pre K-8
Article title: Helping kids think and work on their own
Pages: 27-30
Abstract: Presents several techniques that can boost students' confidence levels, increase independent work habits, and help maximize communication skills on school test day and beyond.
CHECK ONLY ONE ANSWER.
☐ general public
☐ scholars
☐ students
☐ teachers

49. You have been asked to make a presentation to your class on assistive technologies available in the schools for visually challenged students. From which source would you be most likely to find appropriate information?
CHECK ONLY ONE ANSWER.
☐ *Elementary School Journal*
☐ *Information Technology and Disabilities*
☐ Journal of Learning Disabilities
☐ *Teaching Exceptional Children*
☐ *Technology Teacher*

50. The following item was retrieved from an ERIC database search. What kind of source is it?
Title: Pre-service Elementary Teachers' Self-Efficacy Beliefs
Author(s): Cakiroglu, Jale; Boone, William J.
Publication Year: 2001
Abstract: The purpose of this study was to examine pre-service elementary teachers' self-efficacy beliefs in teaching science.
Notes: Presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
Number of Pages: 24
ERIC Number: ED453084
CHECK ONLY ONE ANSWER.
☐ a book
☐ a book chapter
☐ a book review
☐ a conference paper
☐ a journal article

51. You are discussing the controversial book *Cultural Literacy* in your class and want to find out more about the qualifications of the author, E. D. Hirsch. The most objective information may be found by:
CHECK ONLY ONE ANSWER.
□ locating author information in biographical publications
□ locating author information on the Web
□ locating reviews of the book
□ looking at author information in the book

52. While developing a lesson plan on the U.S. legislative system, you find the following story on the Internet:
Congress Launches National Congress-Awareness Week
WASHINGTON, DC—Hoping to counter ignorance of the national legislative body among U.S. citizens, congressional leaders named the first week in August National Congress Awareness Week. "This special week is designed to call attention to America's very important federal lawmaking body," Speaker of the House Dennis Hastert said. The festivities will kick off with a 10-mile Walk for Congress Awareness.
The item is from a newspaper Web site, which states it is "America's Finest News Source." Given this, the following action is in order:
CHECK ONLY ONE ANSWER.
□ although the story was found on the Web, you can use it as it's obviously from a reputable news service
□ you decide to investigate the reputation of the publisher by looking at their Web site
□ you decide to investigate the reputation of the publisher by looking at other Web sites

53. To find general information on "direct instruction," you perform a Google search and retrieve the following results. Which of the following sources may reflect the author's, publisher's, or sponsor's point of view?
CHECK ONLY ONE ANSWER.
□ first page is entitled: Effective literacy instruction for adolescents, commissioned by the National Reading Conference, URL: http://www.nrconline.org/publications/alverwhite2.pdf
□ first page is entitled: This paper examines Direct Instruction--one branch of the "instructivist" approach in education, URL: http://people.uncw.edu/kozloffm/diarticle.pdf
□ first page is entitled: What the data really show: Direct instruction really works! The dirty little secret from the biggest education study ever, URL: http://www.jefflindsay.com/EducData.shtml
□ all of the results
□ none of the results

54. You have been assigned a paper to investigate how group work impacts student learning. A keyword search in ERIC on group work has returned over 600 items. To narrow your search, which of the following steps would you next perform?
CHECK ONLY ONE ANSWER.
☐ add impacts as a keyword
☐ add student learning as a keyword
☐ limit search results by date
☐ limit search results by publication type

55. When one finds an excellent article, which of the following is the most direct method for identifying related sources?
CHECK ONLY ONE ANSWER.
☐ ask a reference librarian
☐ ask the course instructor
☐ look at the bibliography from the article
☐ search ERIC under the author's name
☐ search the library catalog under the author's name

56. When is it ethical to use the ideas of another person in a research paper?
CHECK ONLY ONE ANSWER.
☐ it is never ethical to use someone else's ideas
☐ only if you do not use their exact words
☐ only when you give them credit
☐ only when you receive their permission
☐ only when you use their exact words

57. Which of the following strategies may result in a plagiarism violation?
CHECK ONLY ONE ANSWER.
☐ document your source by citing the information you use
☐ paraphrase the idea so you do not need to cite it
☐ take accurate notes as to where you found specific ideas
☐ for exact quotations, use quotation marks and credit the author

58. Using this result from an Internet search engine, who is the "owner" of this Web site?
State policies on planning, funding, and standards. Does the state have technology requirements for students?
http://www.edweek.org/reports/tc98/states/fl.htm
CHECK ONLY ONE ANSWER.
☐ business or commercial entity
☐ college or university
☐ national government agency
☐ other organization
☐ state government agency

59. Based on the following paragraph, which sentence should be cited?

(1)Technology use in the schools is often characterized as a potentially dehumanizing force. (2)Perhaps the fear that the virtual world may lead to passivity and isolation, at the expense of literal social interaction, is valid. (3)Certainly, educators must ask *which* uses of technology result in increased learning and a better quality of life. (4)To address these issues, one model has been proposed that suggests students work in groups with the computer peripheral to the group and the teacher serves as facilitator.

CHECK ONLY ONE ANSWER.

☐ 1
☐ 2
☐ 3
☐ 4

60. You are creating a Web page for a student education organization. Browsing the Internet, you find a neat photo from the US Department of Education, which is a government agency. If you decide to use the graphic on your Web page, which of the following copyright choices is the proper action?

CHECK ONLY ONE ANSWER.

☐ Permission is not needed as the photo is from a government agency.
☐ Permission is not needed as the photo was found on the Internet.
☐ Permission is not needed as you are only using it for a Web page.
☐ Permission to use the photo must be acquired before using it.

61. You have an assignment that requires you to use course management software to practice setting up a class grade book. Your school has purchased the software and loaded it in the computer lab, but you have a difficult time getting to the lab due to work conflicts. A friend loans you the software and you load it on your computer. Is this legal?
CHECK ONLY ONE ANSWER.
☐ No, because this action constitutes a violation of copyright.
☐ Yes, because it is already freely available in the lab.
☐ Yes, because it is education software and therefore able to be shared.
☐ Yes, because your friend owns it and can share as he wants.

62. Browsing a weekly news magazine, you come across an article that discusses the future of space exploration. As you are teaching this topic you decide to make 25 copies of the article and share it with your class.
Which of the following concepts makes it legally permissible to reproduce portions of works for educational purposes without permission?
CHECK ONLY ONE ANSWER.
☐ Copyright
☐ Fair use
☐ Freedom of information
☐ Intellectual freedom
☐ Intellectual property

APPENDIX G:
BEILE TEST OF INFORMATION LITERACY FOR EDUCATION (B-TILED)

Beile Test of Information Literacy for Education (B-TILED)

The library is gathering information to evaluate the effectiveness of its instruction program. This questionnaire consists of demographic questions and a library and information skills quiz.

Fill in the most correct choice on your Scantron form.

1. Overall, how would you rate your ability to search library databases to find information?
   a. excellent
   b. good
   c. average
   d. poor

2. Overall, how would you rate your ability to search the Internet to find information?
   a. excellent
   b. good
   c. average
   d. poor

Please indicate whether you have attended any of the following since you began your studies at UCF.

3. Have you attended a tour or physical orientation of the library?
   a. yes
   b. no
   c. don't know

4. Have you attended a library instruction session held in your classroom?
   a. yes
   b. no
   c. don't know

5. Have you attended a library instruction session held in the library?
   a. yes
   b. no
   c. don't know

6. Have you had one on one intensive instruction with a librarian?
   a. yes
   b. no
   c. don't know

7. Which of the following characteristics best indicates scholarly research?
   a. available in an academic library
   b. indexed by ERIC
   c. reviewed by experts for publication
   d. written by university faculty

8. Your professor has assigned a paper on the whole language movement. You are not familiar with the topic, so you decide to read a brief history and summary about it. Which of the following sources would be best?
   a. a book on the topic, such as *Perspectives on whole language learning: A case study*
   b. a general encyclopedia, such as *Encyclopedia Britannica*
   c. an article on the topic, such as "Whole language in the classroom: A student teacher's perspective."
   d. an education encyclopedia, such as *Encyclopedia of Education*

9. Research or periodical databases are designed to include items based on which of the following criteria?
   a. found on the Internet
   b. not found on the Internet
   c. owned by your library
   d. relevant subject matter

10. ERIC is the most appropriate database to search to locate:
   a. education article citations and documents
   b. education publications from 1877 to current
   c. full-text education articles
   d. US Department of Education statistics

11. Most research and periodical databases have basic and advanced searching interfaces. Which of the following can you do ONLY in advanced searching?
   a. add Boolean or search connectors between terms
   b. enter multiple search terms
   c. search by keyword
   d. search multiple terms by field

12. Research studies in education are generally first communicated through:
    a. books published by education associations
    b. education encyclopedia entries
    c. newsletters of education associations
    d. professional conferences and journal articles

13. You have been assigned to write a short class paper on effective instruction techniques for teaching English as a Second Language (ESL) students. Your professor indicated three recent scholarly sources would be sufficient. Which strategy is best to locate items?
    a. search a general academic and an education database for journal articles
    b. search an education database for journal articles
    c. search the library catalog for books
    d. search the library catalog for encyclopedias

14. Select the set of search terms that best represent the main concepts in the following:
    What are the health risks associated with the use of drug therapy for hyperactive students?
    a. drug therapy, health risks, hyperactivity
    b. drug therapy, health risks, students
    c. drug therapy, hyperactivity, students
    d. drugs, hyperactivity, therapy

15. Select the set that best represents synonyms and related terms for the concept "college students."
    a. colleges, universities, community colleges…
    b. Gen X, students, undergraduates…
    c. graduate students, freshmen, sophomores...
    d. university, adult learners, educational attendees...

16. While researching a paper on character education, you find that it is also sometimes called *values education* or *moral education*. You decide to look for information on the subject in a research database, and to save time you write a search statement that includes all three terms. Which of the following is the best example to use when you have fairly synonymous terms and it does not matter which of the terms is found in the record?
    a. character and values and moral
    b. character or values or moral
    c. character, values and moral
    d. character, values or moral

17. You are using a research database that uses an asterisk (*) as its truncation symbol. When you type in *read** you would retrieve records that contained which of the following words?
    a. examine, peruse, reader, reading
    b. peruse, read, reader, reading
    c. read, reader, reads, readmit
    d. read, reader, reading, reapply

18. You have a class assignment to investigate how group work impacts student learning. A keyword search in ERIC on "group work" has returned over 600 items. To narrow your search, which of the following steps would you next perform?
    a. add "impacts" as a keyword
    b. add "student learning" as a keyword
    c. limit search results by date
    d. limit search results by publication type

19. The following citation is for:
    Massaro, D. (1991). Broadening the domain of the fuzzy logical model of perception. In H. L. Pick, Jr., P. van den Broek, & D. C. Knill (Eds.), *Cognition: Conceptual and methodological issues* (pp. 51-84). Washington, DC: American Psychological Association.
    a. a book
    b. a chapter in a book
    c. a journal article
    d. an ERIC document

20. Your professor suggested you read a particular article and gave you the following citation:
    Shayer, M. (2003). Not just Piaget, not just Vygotsky. *Learning and Instruction, 13*(5), 465-485.
    Which of the following would you type into the library's catalog to locate the actual article?
    a. author search: Shayer
    b. journal title search: *Learning and Instruction*
    c. journal title search: Not just Piaget, not just Vygotsky
    d. subject search: Piaget and Vygotsky

21. The following item was retrieved from an ERIC database search. What kind of source is it?
Title: Pre-service Elementary Teachers' Self-Efficacy Beliefs
Author(s): Cakiroglu, Jale; Boone, William J.
Publication Year: 2001
Abstract: The purpose of this study was to examine pre-service elementary teachers' self-efficacy beliefs in teaching science.
Notes: Presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
Number of Pages: 24
ERIC Number: ED453084
a. a book
b. a book chapter
c. a conference paper
d. a journal article

22. Using this result from an Internet search engine, who is the "owner" of this Web site?
State policies on planning, funding, and standards. Does the state have technology requirements for students?
http://www.edweek.org/reports/tc98/states/fl.htm
a. business or commercial entity
b. college or university
c. other organization
d. state government agency

23. While developing a lesson plan on the U.S. legislative system, you find the following story on the Internet:
Congress Launches National Congress-Awareness Week
WASHINGTON, DC—Hoping to counter ignorance of the national legislative body among U.S. citizens, congressional leaders named the first week in August National Congress Awareness Week. "This special week is designed to call attention to America's very important federal lawmaking body," Speaker of the House Dennis Hastert said. The festivities will kick off with a 10-mile Walk for Congress Awareness.
The item is from a newspaper Web site, which states it is "America's Finest News Source." Given this, the following action is in order:
a. you can use the story as it's obviously from a reputable news source
b. you decide to investigate the reputation of the publisher by looking at their Web site
c. you decide to investigate the reputation of the publisher by looking at other Web sites
d. you should not use the story because Web information is not always trustworthy

24. Based on the following paragraph, which sentence should be cited?
(1)Technology use in the schools is often characterized as a potentially dehumanizing force. (2)Perhaps the fear that the virtual world may lead to passivity and isolation, at the expense of literal social interaction, is valid. (3)Certainly, educators must ask *which* uses of technology result in increased learning and a better quality of life. (4)To address these issues, Hunter has proposed that students work in groups with the computer peripheral to the group and the teacher acting as facilitator.

a. 1
b. 2
c. 3
d. 4

25. When is it ethical to use the ideas of another person in a research paper?

a. it is never ethical to use someone else's ideas
b. only if you do not use their exact words
c. only when you give them credit
d. only when you receive their permission

26. You are planning an open house for your students' parents. Browsing the Internet, you find the report *Child Safety on the Internet*, which is a US Department of Education publication. If you distribute 30 copies of the report to parents at the open house, which of the following copyright choices is the proper action?

a. permission is not needed as the report is from a government agency.
b. permission is not needed as the report was found on the Internet.
c. permission is not needed as you are only distributing 30 copies.
d. permission to distribute 30 copies of the report must be acquired.

27. You have an assignment that requires you to use course management software to practice setting up a class grade book. Your school has purchased the software and loaded it in the computer lab, but you have a difficult time getting to the lab due to work conflicts. A friend loans you the software and you load it on your computer. Is this legal?

a. no, because this action constitutes a violation of copyright.
b. yes, because it is already freely available in the lab.
c. yes, because it is education software and therefore able to be shared.
d. yes, because your friend owns it and can share as he wants.

28. Browsing a weekly news magazine, you come across an article that discusses the future of space exploration. As you are teaching this topic you decide to make copies of the article and share it with your class.
Which of the following concepts makes it legally permissible to reproduce portions of works for educational purposes without permission?
a. copyright
b. fair use
c. freedom of information
d. intellectual freedom

29. Which of the following most closely describes the level you want to teach?
a. early childhood
b. elementary
c. middle school
d. high school

30. What is your student classification?
a. freshman
b. sophomore
c. junior
d. senior

31. How long have you been continuously enrolled at UCF?
a. less than 1 year
b. 1 to 2 years
c. 3 to 4 years
d. more than 4 years

32. Have you ever attended another university or college?
a. yes (go to question 33)
b. no (skip to question 34)

33. How long ago did you attend another university or college?
a. 0-1 year
b. 2-3 years
c. 4-5 years
d. more than 5 years

34. What is your gender?
a. male
b. female

35. Please indicate those racial or ethnic groups that apply to you. (Select all that apply.)

a. White or European American
b. Hispanic or Latino
c. Black or African American
d. Asian or Asian American
e. Other (write in on Scantron)

Thank you!

Test Key

7. C
8. D
9. D
10. A
11. D
12. D
13. B
14. A
15. C
16. B
17. C
18. B
19. B
20. B
21. C
22. C
23. C
24. D
25. C
26. A
27. A
28. B

APPENDIX H:
THE IN-LIBRARY TEST FOR CRITERION-RELATED VALIDITY

8. [Show student a relevant book, a general encyclopedia, a relevant journal article, and an education encyclopedia.] "Which of these is the best source if you want to read a brief history and summary of the whole language movement?"

12. [Show student an education association-published book, an education encyclopedia, an education association newsletter, and a journal article.] "Generally, which of these sources is the first to publish research studies in education?"

13. [Seat student at a workstation showing the library web page.] "You need three recent scholarly articles on teaching English as a second language. What is your next step?"

18. [Show student *ERIC* database screen, with search strategy showing the results from a keyword search on *group work*.] "You are searching how group work impacts student learning. What is the next logical step when you get this many results?"

20. [Show student library catalog search screen. Hand student following citation: Shayer, M. (2003). Not just Piaget, not just Vygotsky. *Learning and Instruction, 13*(5), 465-485.]. "Type in what you need to locate the item."

22. [Show student Internet screen with the following URL entered: http://www.edweek.org/reports/tc98/states/fl.htm.] "Who owns or sponsors this site?"

24. [Hand student the following paragraph. Technology use in the schools is often characterized as a potentially dehumanizing force. Perhaps the fear that the virtual world may lead to passivity and isolation, at the expense of literal social interaction, is valid. Certainly, educators must ask *which* uses of technology result in increased learning and a better quality of life. To address these issues, Hunter has proposed that students work in groups with the computer peripheral to the group and the teacher acting as facilitator.] "Which sentence should be cited?"

26. [Hand student a copy of the government publication *Youth, Pornography, and the Internet*.] "If you found this government document on the Internet and wanted to hand out 30 copies to your students' parents at an open house, what is the appropriate copyright action?"

APPENDIX I:
STUDY TIMELINE

Study Timeline

January 2004	Submitted a fellowship proposal, with Gordon Taub, to Project SAILS to develop education-specific information literacy test items.
February 2004	Received notification of award.
March 2004	Reviewed materials sent by Project SAILS.
April 2004	Attended Project SAILS week-long training institute at Kent State with three other project teams. Submitted work plan to Project SAILS.
May 2004	Had work plan approved. Identified and contacted content experts. Reviewed standards for test content, submitted document to Project SAILS in mid-May. Commenced item writing immediately.
June 2004	Had test content parameters approved. Continued items writing for content areas. Submitted IRB human subjects approval request. Submitted draft items to Project SAILS and content experts.
July 2004	Received IRB approval. Conducted individual testing with 6 students. Item comments received from Project SAILS and content experts.
August 2004	Revised and formatted items for survey based on feedback from Project SAILS, students, and reviewers. Submitted item revisions to Project SAILS.
September 2004	Conducted small group testing (around hurricanes). Entered results and ran item analysis procedures. Submitted final report to Project SAILS.
October 2004	Spoke with OEAS regarding test administration. Presented proposal to develop instrument for dissertation project.. Started reducing items from 62 to 22 for test.
November 2004	Submitted survey and informed consent form to OEAS. Invitation to participate in study sent to education majors by OEAS.
December 2004	Data file received from OEAS. Data reviewed and "cleaned." Commenced planning for supplemental test administration.
January 2005	Discussed purpose and status of study at ALA Conference.
February 2005	Administered 80 more surveys over 6 testing days. Continued inputting data. Performed criterion-related validity procedures. Performed test stability procedures.
March 2005	Ran statistical analyses. Began writing up results. Presented poster session at UCF GSA Research Forum.
April 2005	Presented findings at AERA.
May-Aug 2005	Continued writing dissertation. Submitted dissertation draft to committee.

APPENDIX J:
CONTENT REVIEWER ITEM SUMMARY GRID

ITEM ACCURACY RATING

	Item	1	2	3	4	5
Reviewer #1		2	3	3	2	3
Reviewer #2		2	3	3	1	3
Reviewer #3		3	2	3	2	3
Reviewer #4		3	3	0	2	3
Reviewer #5		3	2	1	1	2
		13	13	10	8	14
Mean		2.6	2.6	2	1.6	2.8

	Item	6	7	8	9	10
Reviewer #1		3	1	3	3	3
Reviewer #2		3	2	3	2	3
Reviewer #3		2	3	3	1	2
Reviewer #4		3	0	2	0	1
Reviewer #5		3	3	2	3	3
		14	9	13	9	12
Mean		2.8	1.8	2.6	1.8	2.4

	Item	11	12	13	14	15
Reviewer #1		3	3	3	3	3
Reviewer #2		3	2	3	3	3
Reviewer #3		3	2	3	2	3
Reviewer #4		2	3	3	2	3
Reviewer #5		3	3	3	2	3
		14	13	15	12	15
Mean		2.8	2.6	3	2.4	3

	Item	16	17	18	19	20
Reviewer #1		3	3	3	3	3
Reviewer #2		2	2	2	2	3
Reviewer #3		3	1	2	3	3
Reviewer #4		3	3	2	3	3
Reviewer #5		1	3	3	3	3
		12	12	12	14	15
Mean		2.4	2.4	2.4	2.8	3

	Item	21	22	23	24	25
Reviewer #1		2	3	3	3	3
Reviewer #2		3	3	3	3	1
Reviewer #3		2	3	3	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	3	3	3	3
		13	15	15	15	13
Mean		2.6	3	3	3	2.6

	Item	26	27	28	29	30
Reviewer #1		3	3	3	2	3
Reviewer #2		3	3	3	2	2
Reviewer #3		3	3	3	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	3	3	3	3
		15	15	15	13	14
Mean		3	3	3	2.6	2.8

	Item	31	32	33	34	35
Reviewer #1		3	3	1	2	3
Reviewer #2		2	2	2	2	3
Reviewer #3		3	3	2	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	3	3	2	3
		14	14	11	12	15
Mean		2.8	2.8	2.2	2.4	3

	Item	36	37	38	39	40
Reviewer #1		3	3	3	3	3
Reviewer #2		3	2	3	2	2
Reviewer #3		3	3	3	3	3
Reviewer #4		3	3	3	2	2
Reviewer #5		3	3	3	3	3
		15	14	15	13	13
Mean		3	2.8	3	2.6	2.6

	Item	41	42	43	44	45
Reviewer #1		2	3	3	3	2
Reviewer #2		2	3	3	3	3
Reviewer #3		3	3	3	3	2
Reviewer #4		2	3	3	3	3
Reviewer #5		3	3	3	3	3
		12	15	15	15	13
Mean		2.4	3	3	3	2.6

	Item	46	47	48	49	50
Reviewer #1		3	3	3	3	3
Reviewer #2		2	2	3	1	3
Reviewer #3		2	3	2	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	3	3	3	3
		13	14	14	13	15
Mean		2.6	2.8	2.8	2.6	3

	Item	51	52	53	54	55
Reviewer #1		3	3	1	3	3
Reviewer #2		3	3	3	2	3
Reviewer #3		3	3	3	3	3
Reviewer #4		3	3	2	3	3
Reviewer #5		3	2	3	3	3
		15	14	12	14	15
Mean		3	2.8	2.4	2.8	3

	Item	56	57	58	59	60
Reviewer #1		3	3	3	3	2
Reviewer #2		2	2	2	3	2
Reviewer #3		3	3	3	2	2
Reviewer #4		3	3	1	2	2
Reviewer #5		3	3	2	2	3
		14	14	11	12	11
Mean		2.8	2.8	2.2	2.4	2.2

	Item	61	62
Reviewer #1		3	2
Reviewer #2		3	2
Reviewer #3		2	2
Reviewer #4		2	2
Reviewer #5		3	3
		13	11
Mean		2.6	2.2

ITEM CLARITY RATING

	Item	1	2	3	4	5
Reviewer #1		3	3	1	2	3
Reviewer #2		2	2	2	1	3
Reviewer #3		3	3	3	2	3
Reviewer #4		3	3	0	1	3
Reviewer #5		3	2	2	2	3
		14	13	8	8	15
Mean		2.8	2.6	1.6	1.6	3

	Item	6	7	8	9	10
Reviewer #1		2	1	3	3	3
Reviewer #2		2	2	2	1	2
Reviewer #3		2	2	3	1	2
Reviewer #4		3	0	3	0	1
Reviewer #5		3	3	3	2	3
		12	8	14	7	11
Mean		2.4	1.6	2.8	1.4	2.2

	Item	11	12	13	14	15
Reviewer #1		3	3	3	3	3
Reviewer #2		3	2	3	3	3
Reviewer #3		2	2	3	2	3
Reviewer #4		1	3	3	1	3
Reviewer #5		3	2	3	3	3
		12	12	15	12	15
Mean		2.4	2.4	3	2.4	3

	Item	16	17	18	19	20
Reviewer #1		3	3	3	1	3
Reviewer #2		3	1	2	3	3
Reviewer #3		3	3	2	3	3
Reviewer #4		3	3	0	2	3
Reviewer #5		3	3	3	3	2
		15	13	10	12	14
Mean		3	2.6	2	2.4	2.8

	Item	21	22	23	24	25
Reviewer #1		2	3	2	2	3
Reviewer #2		3	1	3	2	2
Reviewer #3		2	3	3	3	3
Reviewer #4		3	3	3	2	3
Reviewer #5		3	3	2	2	3
		13	13	13	11	14
Mean		2.6	2.6	2.6	2.2	2.8

	Item	26	27	28	29	30
Reviewer #1		3	3	2	2	3
Reviewer #2		3	1	1	1	3
Reviewer #3		3	3	3	2	3
Reviewer #4		2	2	1	1	3
Reviewer #5		3	3	3	2	3
		14	12	10	8	15
Mean		2.8	2.4	2	1.6	3

	Item	31	32	33	34	35
Reviewer #1		3	2	2	3	3
Reviewer #2		3	3	3	3	3
Reviewer #3		3	3	2	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	3	3	2	3
		15	14	13	14	15
Mean		3.0	2.8	2.6	2.8	3.0

	Item	36	37	38	39	40
Reviewer #1		3	3	3	1	2
Reviewer #2		3	2	2	2	2
Reviewer #3		3	3	3	3	3
Reviewer #4		3	3	3	2	3
Reviewer #5		3	2	3	3	3
		15	13	14	11	13
Mean		3.0	2.6	2.8	2.2	2.6

	Item	41	42	43	44	45
Reviewer #1		3	2	2	2	3
Reviewer #2		3	2	2	2	3
Reviewer #3		2	3	3	3	2
Reviewer #4		1	3	3	3	3
Reviewer #5		3	3	3	3	2
		12	13	13	13	13
Mean		2.4	2.6	2.6	2.6	2.6

	Item	46	47	48	49	50
Reviewer #1		1	2	2	3	3
Reviewer #2		3	3	3	3	3
Reviewer #3		2	3	2	3	3
Reviewer #4		3	3	2	3	3
Reviewer #5		2	3	3	3	3
		11	14	12	15	15
Mean		2.2	2.8	2.4	3	3

	Item	51	52	53	54	55
Reviewer #1		2	3	2	3	1
Reviewer #2		2	2	1	2	2
Reviewer #3		3	3	3	3	3
Reviewer #4		2	1	3	3	3
Reviewer #5		3	2	1	3	3
		12	11	10	14	12
Mean		2.4	2.2	2	2.8	2.4

	Item	56	57	58	59	60
Reviewer #1		2	1	1	1	2
Reviewer #2		2	1	3	2	3
Reviewer #3		3	2	3	2	3
Reviewer #4		3	1	2	2	2
Reviewer #5		3	3	1	2	3
		13	8	10	9	13
Mean		2.6	1.6	2	1.8	2.6

	Item	61	62
Reviewer #1		1	2
Reviewer #2		2	3
Reviewer #3		2	2
Reviewer #4		2	3
Reviewer #5		1	2
		8	12
Mean		1.6	2.4

INSTITUTIONAL OBJECTIVITY RATING

	Item	1	2	3	4	5
Reviewer #1		3	3	3	3	3
Reviewer #2		3	3	2	3	3
Reviewer #3		3	3	3	3	3
Reviewer #4		3	3	2	3	3
Reviewer #5		3	3	3	3	3
		15	15	13	15	15
Mean		3	3	2.6	3	3

	Item	6	7	8	9	10
Reviewer #1		3	2	3	3	3
Reviewer #2		3	1	3	2	3
Reviewer #3		2	3	3	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	2	3	3	3
		14	11	15	14	15
Mean		2.8	2.8	3	2.8	3

	Item	11	12	13	14	15
Reviewer #1		3	1	3	3	3
Reviewer #2		2	1	3	3	3
Reviewer #3		3	2	3	3	3
Reviewer #4		3	1	3	3	3
Reviewer #5		0	2	3	3	3
		11	7	15	15	15
Mean		2.2	1.4	3	3	3

	Item	16	17	18	19	20
Reviewer #1		3	3	3	3	3
Reviewer #2		2	2	3	3	3
Reviewer #3		3	3	2	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	3	3	3	3
		14	14	14	15	15
Mean		2.8	2.8	2.8	3	3

	Item	21	22	23	24	25
Reviewer #1		3	3	3	3	3
Reviewer #2		1	3	3	2	1
Reviewer #3		2	3	3	2	2
Reviewer #4		3	3	3	1	3
Reviewer #5		3	3	3	3	3
		12	15	15	11	12
Mean		2.4	3	3	2.2	2.4

	Item	26	27	28	29	30
Reviewer #1		3	3	3	3	3
Reviewer #2		3	3	1	3	1
Reviewer #3		2	2	3	3	2
Reviewer #4		2	3	1	3	2
Reviewer #5		3	3	3	3	3
		13	14	11	15	11
Mean		2.6	2.8	2.2	3	2.2

	Item	31	32	33	34	35
Reviewer #1		2	2	3	3	3
Reviewer #2		1	1	3	3	1
Reviewer #3		3	3	3	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	3	3	3	3
		12	12	15	15	13
Mean		2.4	2.4	3	3	2.6

	Item	36	37	38	39	40
Reviewer #1		3	3	3	3	3
Reviewer #2		3	2	1	2	2
Reviewer #3		3	3	3	3	3
Reviewer #4		3	2	3	3	3
Reviewer #5		3	3	3	3	3
		15	13	13	14	14
Mean		3	2.6	2.6	2.8	2.8

	Item	41	42	43	44	45
Reviewer #1		2	3	3	3	3
Reviewer #2		2	3	3	3	3
Reviewer #3		3	3	3	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	3	3	3	3
		13	15	15	15	15
Mean		2.6	3	3	3	3

	Item	46	47	48	49	50
Reviewer #1		3	3	3	3	3
Reviewer #2		3	2	3	3	3
Reviewer #3		3	3	3	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	3	3	3	3
		15	14	15	15	15
Mean		3	2.8	3	3	3

	Item	51	52	53	54	55
Reviewer #1		3	3	3	3	3
Reviewer #2		3	3	3	3	3
Reviewer #3		3	3	3	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	3	3	3	3
		15	15	15	15	15
Mean		3	3	3	3	3

	Item	56	57	58	59	60
Reviewer #1		3	3	3	3	3
Reviewer #2		3	3	3	3	3
Reviewer #3		3	3	3	3	3
Reviewer #4		3	3	3	3	3
Reviewer #5		3	3	3	3	3
		15	15	15	15	15
Mean		3	3	3	3	3

	Item	61	62
Reviewer #1		3	3
Reviewer #2		3	3
Reviewer #3		3	3
Reviewer #4		3	3
Reviewer #5		3	3
		15	15
Mean		3	3

APPENDIX K:
CURRICULUM VITAE

PENNY BEILE O'NEIL
Associate Librarian

University of Central Florida
UCF Libraries
Curriculum Materials Center
PO Box 162666
Orlando, FL 32816-2666

Phone: (W) 407-823-5488
(H) 407-568-6831
Email: pbeile@mail.ucf.edu
Web sites: http://library.ucf.edu/cmc
http://pegasus.cc.ucf.edu/~pbeile

EDUCATION

Ph.D., Curriculum and Instruction, University of Central Florida, Orlando, FL, 2005.
Dissertation topic: *Development and Validation of the Beile Information Literacy Test for Education (B-TILED)*

M.S., Library Science, University of Kentucky, Lexington, KY, 1991.

M.A., Education, University of Kentucky, Lexington, KY, 1989.

B.A., Anthropology and Political Science, dual major, University of Kentucky, Lexington, KY, 1987.

PROFESSIONAL EXPERIENCE

Associate Librarian, Head, Curriculum Materials Center, University of Central Florida Libraries, University of Central Florida, Orlando, FL, June 2000-present.

Assistant Librarian, Reference Librarian, University of Central Florida Libraries, University of Central Florida, Orlando, FL, June 1998-June 2000.

Assistant Librarian, Head, Education Resources, Louisiana State University Libraries, Louisiana State University, Baton Rouge, LA, August 1995-June 1998.

Assistant Librarian, Social Sciences Reference, Louisiana State University Libraries, Louisiana State University, Baton Rouge, LA, January 1994-June 1998.

Librarian (non-rank), **Social Sciences Reference**, Miami University Libraries, Miami University, Oxford, OH, January 1992-January 1994.

OTHER EXPERIENCE

Fugazzi College, Lexington, Kentucky:
Librarian (non-rank), January 1991-December 1991.

University of Kentucky, Lexington, Kentucky:
School of Library and Information Science, Graduate Assistant, August 1990-August 1991.
College of Education, Graduate Teaching Assistant, August 1987-May 1988.

Items with an asterisk denote activities or publications that have taken place since my dossier for promotion to Associate Librarian rank was submitted in August, 2000.

1. PUBLICATIONS

A. Published

Books and chapters in books:

* Beile, Penny M. (2004). "The Development and Use of a Web-based Library Tutorial for Education Students," in Patricia Libutti (Ed.), *Digital Resources and Education Libraries: Innovation, Invention, and Implementation*. ACRL: Chicago.

Articles in refereed journals, refereed bulletins, and refereed proceedings:

* Boote, David N. and Penny M. Beile. "Scholars before Researchers: On the Centrality of the Dissertation Literature Review in Research Preparation." *Educational Researcher* 34(6), (August/September 2005): 3-15 [Featured Article].

* Beile, Penny M., David N. Boote, and Elizabeth K. Killingsworth. "A Microscope or a Mirror?: A Question of Study Validity Regarding the Use of Dissertation Citation Analysis for Evaluating Research Collections (in Education)." *Journal of Academic Librarianship* (30)5, (September 2004): 347-353.

* Beile, Penny M. "Effectiveness of Course-Integrated and Repeated Library Instruction on Library Skills of Education Students." *Journal of Educational Media and Library Sciences* 40(3), (March 2003): 271-277.

* Beile, Penny M. and David N. Boote. "Library Instruction and Graduate Professional Development: Exploring the Effect of Learning Environments on Self-efficacy and Learning Outcomes." *Alberta Journal of Educational Research* 68(4), (Winter 2002): 364-367 [Research Note].

Beile, Penny M. and Megan M. Adams. "Other Duties as Assigned: Emerging Trends in the Academic Library Job Market." *College and Research Libraries* 61, (July 2000): 336-347.

Beile, Penny M. "Great Expectations: Competency-Based Training for Student Media Center Assistants." *MC Journal: The Journal of Academic Media Librarianship* 5(2), (1997). URL: http://wings.buffalo.edu/publications/mcjrnl/v5n2/

Other:

Atkins, David and Penny M. Beile. *Louisiana Academic Library Statistics 1994: Selected Statistics from the 1994 IPEDS Survey of Academic Libraries*. Baton Rouge: LSU Libraries, 1995. Publication distributed to all Louisiana academic libraries. Published by Resources in Education, Education Resources Information Center database, November 1995 (ED384361).

(4) Book reviews:

Review of Brier, Bob. *The Encyclopedia of Mummies*. New York: Facts on File, 1998. In *Choice* 35 (July 1998): 1825.

Review of CD-ROM version of *Columbia Encyclopedia*. New York: Columbia University Press, 1997. In *Choice* 35 (Oct 1997): 263.

Review of Hellebust, Lynn. *State Legislative Sourcebook 1996*. Topeka, KS: Government Research Service, 1996. In *Choice* 33 (Sept 1996): 99.

With Thomas O'Neil, review of CD-ROM version of *Beacham's International Threatened, Endangered, and Extinct Species*. Osprey, FL: Beacham Publishing Co., 1995. In *Choice* 33 (Jan 1996): 77

Review of Walters, Gregory J. *Human Rights in Theory and Practice: A Selected and Annotated Bibliography*. Scarecrow/Salem Press, 1995. In *Choice* 33 (Jan 1996): 768.

Review of Baxter, Pam M. *Psychology: A Guide to Reference and Information Sources*. Englewood, CO: Libraries Unlimited, 1993. In *Research Strategies* 14 (Winter 1995): 64.

Review of Bennett, James R. *Political Prisoners and Trials: A Worldwide Annotated Bibliography, 1900 through 1993*. McFarland, 1995. In *Choice* 32 (Nov 1995): 429.

Review of Chan, Anja Angelica. *Women and Sexual Harassment: A Practical Guide to the Legal Protections of Title VII and the Hostile Environment Claim*. New York: Harrington Park Press, 1994. In *Social Responsibilities Round Table Newsletter* #116 (June 1995): 13.

Review of the eighth edition of *Worldmark Encyclopedia of the Nations*. Detroit: Gale Research, 1995. In *Choice* 32 (June 1995): 1581.

Review of the Oxford University Press edition of *Encyclopedic World Atlas*. London: Oxford University Press, 1994. In *Choice* 32 (March 1995): 1082.

Accepted for publication:

* Beile, Penny M. and David N. Boote. In press for Fall 2005 publication. "Does the Medium Matter?: A Comparison of a Web-based Tutorial with Face-to-Face Library Instruction with Education Students," *Research Strategies,* vol. 19(4).

Submitted for publication:

* Boote, David N. and Penny M. Beile. "The Quality of Dissertation Literature Reviews: A Missing Link in Research Preparation?" Manuscript submitted to *Research in Higher Education*, Summer 2005.

Acknowledgements:

Boster, James S. "The Information Economy Model Applied to Biological Similarity Judgment." In *Perspectives on Socially Shared Cognition*, edited by Lauren B. Resnick, John M. Levine, and Stephanie D. Teasley. Washington, D.C.: APA, 1991. (Note: My contribution to this publication was in the capacity of research assistant for the Kentucky bird study. This is recognized by the author in the acknowledgements.)

2. OTHER SCHOLARLY/RESEARCH/CREATIVE CONTRIBUTIONS

Papers Presented at Professional Meetings

* "Development and Validation of a Standards-based Instrument for Assessing Pre-service Teachers' Information Literacy Levels," delivered at the American Educational Research Association Annual Conference, Montreal, CN, April 2005.
Also presented at the UCF Graduate Research Forum, Orlando, FL, March, 2005.

* "The Quality of Dissertation Literature Reviews: A Missing Link in Professional Preparation," with David N. Boote, delivered (by David Boote) at the American Educational Research Association Annual Conference, San Diego, CA, April 2004.

* "Characteristics of Education Doctoral Dissertation References: Results of an Analysis of Dissertation Citations from Three Institutions," with David N. Boote and Elizabeth K. Killingsworth, delivered at the American Educational Research Association Annual Conference, Chicago, IL, April 2003.
Also presented at the UCF Graduate Research Forum, Orlando, FL, March, 2004.
Also published as *Characteristics of Education Doctoral Dissertation References: An Inter-Institutional Analysis of Review of Literature Citations,* by Resources in Education, Education Resources Information Center, March 2004 (ED478598).

* "The Effect of Library Instruction Learning Environments on Self-Efficacy Levels and Learning Outcomes for Students of Education," with David Boote, delivered at the American Educational Research Association Annual Conference, New Orleans, LA, April 2002.
Also published as *The Effect of Library Instruction Learning Environments on Self-Efficacy Levels and Learning Outcomes for Students of Education,* by Resources in Education, Education Resources Information Center database, April 2002 (ED465331).

"Get Hired!: How to Get the Job You Want," with Megan M. Adams, delivered at the Louisiana Library Association Annual Conference, Shreveport, LA, March 1998.

"Teaching and Learning at a Distance," delivered at the Louisiana Board of Regents Distance Education Initiative, Baton Rouge, LA, February 1998.

"The 1995 LSU Reference Survey and National Level Statistics Gathering Efforts," with David Atkins, delivered at the Louisiana Library Association Annual Conference, Alexandria, LA, March 1996.

Presentations at Professional Meetings, other than Papers

* "Building a Discipline-Specific Web Tutorial," with Jacqueline Toce, presented at the Annual Conference of the American Library Association as a poster session, Atlanta, GA, June 2002.

"Underpaid? Under-employed? Underappreciated?: The Systems Approach to Job Hunting," with Megan Adams, presented at the Annual Conference of the American Library Association as a poster session, San Francisco, CA, June 1997.

"Building Conference Web Sites: A Virtual Tour of Lexington, KY," with Steven Harris, presented at the Biennial Conference of Southeastern Library Association as a poster session, Lexington, KY, October 1996, virtual tour used as the official conference site.

Exhibitions Curated

Miami University:

"Censorship in the 1990's: Banned Books Display," exhibited at King Library and McGuffey Hall (College of Education), August-October 1992.

"Pictorial History of Women at Miami University," co-curator with Jenny Presnell, exhibited at King Library, February-April 1992.

Grants and Fellowships

University of Central Florida:

* Competitive fellowship awarded to UCF Project Team Penny Beile and Gordon Taub for "The Development and Validation of an Education-Specific Information Literacy Assessment Instrument." Fellowship awarded in the amount of $5,000 per team by the Project for the Standardized Assessment of Information Literacy Skills (Project SAILS); funded by the Institute of Museum and Library Services and the Institute for Library and Information Literacy Education.

Louisiana State University:

"The 21st Century Community Learning Center," a combined public and school library initiative for the community of Belle Rose, Assumption Parish, Louisiana, with Janice Stuhlmann Hinson and Alma Dawson, funded in the amount of $1.6 million by the U.S. Department of Education's 21st Century/I Can program, June 2000.

"Acquisition of Hardware to Support Interactive Guide to LSU Libraries," teaching incentive grant, with Margo Brault, funded in the amount of $2,000 by the Office of Academic Affairs and the Center for Faculty Development, January 1995.

Honoraries and Awards

Honoraries:

* Phi Kappa Phi, general academic honor society, 2004-present.
* Pi Lambda Theta, education honor society, 2003-present.
* Kappa Delta Pi, education honor society, 2002-present.
  Beta Phi Mu, library science honor society, 1991-present.
  Phi Beta Kappa, liberal arts honor society, 1985-present.

Awards:

University of Central Florida:

* Excellence in Librarianship award, $1,000 stipend, 2003.

* Feria Internacional, Guadalajara, Mexico travel stipend. $500 awarded by the American Library Association-Federacion de Internacional Librarios Free Pass Program, Nov 2000.

Louisiana State University:

Research stipend awarded for research assistant for content analysis of position descriptions from 1996, Library Faculty Research Award funded in the amount of $300 by the Friends of the LSU Libraries, Spring 1997.

Research stipend awarded for on-site visits to selected instructional materials and media centers in Southern ARL libraries, Library Faculty Research Award funded in the amount of $300 by the Friends of the LSU Libraries, Spring 1996.

Research stipend awarded for data collection for conference Web site for Southeastern Library Association 1996, Library Faculty Research Award funded in the amount of $300 by the Friends of the LSU Libraries, Spring 1995.

Activities Planned or in Progress

Research in progress includes development and validation of an information literacy assessment instrument for teacher education students and analysis of doctoral students in education use of the scholarly literature. Planned future studies include analyzing trends in the academic library job market and their impact on library school curricula, and a critical investigation of institutional representation in top ranked education journals.

3. PERFORMANCE OF PROFESSIONAL RESPONSIBILITIES

Summary of duties at UCF Libraries:

I initially served as a reference librarian with library instruction and collection development responsibilities (June 1998-February 2000). Beginning June 2000, I became head of the Curriculum Materials Center. Administration of the unit entails ensuring adequate staff and student assistant coverage, providing staff development opportunities, maintaining a

comprehensive collection of P-12 teaching materials, leading instruction classes, and promoting the unit and its services to faculty and students.

Significant contributions include:

- Hosting Myrtle Harris, exchange librarian from Jamaica; meeting with and extending services to 11 teacher education students from Aruba; and discussing the role of curriculum materials and associated libraries with representatives from the Sri Lanka Ministry of Education.
- Consulting at a number of local area schools, including Evan Middle School and Nap Ford Charter School.
- Participating in continuous assessment and quality improvement by administering biennial patron surveys and integrating suggestions into training and service.
- With Jackie Toce, developing an extensive Web-based tutorial for distance education students. Assessment of the tutorial revealed it to be comparable to face to face instruction regarding learning outcomes for students.
- Creating student assistant module on Dewey Decimal Classification and shelving using Authorware software. Designing an interactive Jeopardy-style program to quiz students on content covered during CMC orientations.
- Expanding the standardized test collection, merging two collections of tests, and creating a policy for purchase, housing, and circulation.
- Collaborating with personnel from the College of Education and Orange County Public Schools to host several students from the "Transition to Work" program.
Facilitating the donation of a unique collection of oral histories of Central Florida educators conducted by graduate students in Education and a substantial collection of math and science textbooks.
- Strengthening Spanish-language education and picture book collections by attending the International Book Fair and seeking to continue to receive award winning Spanish-language children's books.
- Developing an in-depth CMC library instruction session and marketing it to pre-professional teacher education classes, significantly increasing the number of students receiving instruction. --
- Presenting "Use of manipulatives to facilitate developmentally appropriate instruction" to selected graduate and doctoral level classes.
- Facilitating an upgrade of low-end technology production lab equipment by soliciting production computers and software from a College of Education grant, and requesting additional PCs, laptops, DVD players, scanners, and a digital camera and camcorder from Library Systems.
- Extending services to patrons by circulating the software collection, honoring the Special Borrower Card in the CMC, and initiating a change in library policy whereby Florida educators can freely use the library and CMC collections.

Summary of duties at LSU Libraries:

Duties at LSU included providing in-depth reference and information service and participating in library instruction and outreach programs. I also managed Psychology, Speech Communication, Communication Science and Disorders, Curriculum and Instruction, Administrative and Foundational Services, and Vocational Education library collections, which included creating print and electronic finding aids, acting as liaison to faculty and students in assigned departments, and teaching all discipline library instruction classes.

Significant contributions:

- From June 1994 through August 1995, I served as head of the Reference and Information Desk Work Group.
- From August 1995 to June 1998, I was a member of the Reference Instruction Work Group, where funds from the Student Government Association were solicited to instruct students on how to use "TIGER," a student email and Internet access service.
From June 1994 to June 1998, I acted as facilitator for Social Sciences Collection Development group.
- From July 1997 to June 1998, I assumed an interim position on the Collection Development Management Team. During this time I redesigned YBP approval profile, participated in the serials redesign project, and completed six collection development policies and an extensive policy for Education Resources.
- A significant portion of my time was also spent overseeing the administration of Education Resources, a curriculum materials lab.

4. UNIVERSITY SERVICE

University of Central Florida:

* SACS Quality Enhancement Plan
  Assessment Team, 2005-
* Faculty Senate, Library Senator (elected), 2000-2002.
* Library Advisory Committee, ex-officio, 2000-2002.
* Graduate Council, member, 2000-2002.
* Course Review and New Programs, member, 2000-2002.

University Excellence in Undergraduate Teaching and Advising Awards Committee, reviewer, 1999.

Faculty Center for Teaching and Learning

* Summer Institute, Information Fluency strand, attended at request, 2005.

Summer Institute, library representative, 1999.

Louisiana State University:

Division of Instructional Support and Development, 1996-1998:

Advisory Council, member, 1996-1998.
Most Effective Presentations at the Teaching in Higher Education Forum Review Committee, reviewer, 1998.
Teaching Incentive Grants, reviewer, 1997.

Faculty Senate, 1996-1998:

Committee for the Improvement of Instruction, member, 1996-1998.
Ad Hoc Committee on Writing Across the Curriculum, member, 1996-1998.

Library Service

University of Central Florida:

* Performance Evaluation Committee, 2005.
* ALEPH Implementation and Migration Team, 2005-2006.
  Indexing Team, 2005-2006.
* Quality Enhancement Plan, proposal development member, 2004-2005.
* Mentoring Committee, 2004-2007.
  Chair, 2004-2005.
* Display: *Freedom Fighting: African-Americans in the Military*, 2005.
* Frankenstein exhibit, docent, 2004.

Search Committees:

* AD for Technical Services, 2005.
* Archivist, member, 2003.
* Collection Development Evaluation Criteria Taskforce, member, 2003.
* Copyright Taskforce, member, 2003-2004.
* Cyberterrorism Taskforce, member, 2003.
* Technology Advisory Group, member, 2001-2003.
* Library Dissertation Taskforce, to establish dissertation ILL loan policy, member, 2001.

Electronic Resources Group, to establish procedures for adding electronic items, member, 1999-2000.
Promotion Task Force, to review criteria for promotion, member, 1998-1999.
Reference Editorial Committee, to update reference publications, member, 1998-2000.
Co-coordinator, 1998-1999.
UCF Libraries Faculty Day Committee, to hold a library open house, member, 1998.

Louisiana State University:

Taskforce on Reviewing PS-36 (Faculty Tenure), member, 1997.
Bibliographic Procedures Committee, Ad Hoc Vertical File, member, 1997.
Distance Education Committee, member, 1997-1998.
Library Staff Association (elected), member, 1997-1998.
Vice-president, 1997-1998.
Schwing Lecture Series Committee, member, 1997-1998.
Search Committees:
Science Reference Librarian, chair, 1996.
Martin Luther King, Jr. Committee, reporter, 1996.
Introduction of keynote speaker Dr. Daniel Devore, January 1996.
Taskforce on Designing CARL Gateway, member, 1995.
Library Faculty Policy Committee (elected), member, 1994-1996.
Secretary, 1994-1995.

.

Miami University:

Taskforce on Teaching Portfolios, member, 1993.
Search Committees
Electronic User Education Librarian, member, 1993.
Minority Resident/Librarian, member, 1992.

Taskforce on Designing WWW Front End, member, 1992-1993.
Taskforce for Beta Testing of FirstSearch Database, member, 1992-1993.

Other Service

University of Central Florida:

College of Education
* LiveText Implementation Committee, member, 2003-2004.
* Innovation and Renovation Committee, member, 2000-2003.
* Instructional Resources Committee, ex-officio, 2000-present.
* UCF Education Doctoral Students Association, founding member, 2002-present.
* Doctoral Student Colloquia, with Stacy DeZutter and Nehemiah Ichilov, hosted a four-part series of discussion sessions between doctoral students and faculty.
* "Designing Effective Displays" presented, with Nicole Sotak, to the UCF Student Chapter of the Florida Education Association (SFEA), October 2004.
* "Bulletin Boards that Teach" presented to the Association of Childhood Educators International (ACEI), UCF Student Chapter, October 2000 and October 2003.
* Graduate Student Association, member, 2002-present
* Elections Committee, member, 2003, 2004.
* Graduate Research Forum, volunteer, 2004, moderator, 2005.
* Graduate Student Orientations, volunteer, 2003-2004.

Louisiana State University:

"Get Hired! How to Get the Job You Want," with Aimee Fifarek, presented to the Louisiana State University American Library Association Student Chapter, May 1998.
"Creating a Resume and Enhancing your Interview Skills," with Megan M. Adams, presented to the Louisiana State University American Library Association Student Chapter, March 1996 and April 1997.
"Strategies for Enhancing your Employment Opportunities," two-part workshop, with Mike Somers, presented to the Louisiana State University American Library Association Student Chapter, September 1995 and November 1995.

Miami University:

"In a Nutshell," *Baker Street Gazette*. Five columns presenting biographies of Miami University Libraries employees appearing in issues April 9, April 23, June 25, September 10, and November 19, 1993.
"Electronic Resources for Gerontological Research and Assisted Living Design," brown bag presentation to Scripps Gerontology Center, 1993.
Faculty liaison to Students for the Ethical Treatment of Animals, 1992-1993.
Co-hosted the Friends of the Library Annual Book Sale, 1992.

5. SERVICE TO THE PROFESSION

Professional Memberships

* American Educational Research Association, 2001-present.
* Post-secondary Education Division, 2001-present.
* Communication of Research, Special Interest Group, 2002-present.

American Library Association, 1991-present.
Roundtables:
* Library Research Round Table, 2003-present.
Intellectual Freedom Round Table, 1991-1995.
New Members Round Table, 1996-1998.
Division and Sectionss:
Reference and Adult Service Division, 1993-1996.
Machine Assisted Reference Services Section, 1994-1996.
Association of College and Research Libraries, 1996-present.
Education, Behavioral, and Social Sciences Section, 1996-present.

Southeastern Library Association, 1994-present.
Reference and Adult Services, 1994-1996.
Library Instruction Round Table, 1994-1996.

Florida Library Association, 1998-2000.
Academic Libraries Section, 1998-2000.

Louisiana Library Association, 1994-1998.
Academic Libraries Section, 1994-1998.

Ohio Library Association, 1992-1993.

Service to the Profession

American Library Association
Association of College and Research Libraries Division
Education and Behavioral Sciences Section
* Advisory Council, invited member, 1998-2005.
* Executive Committee, invited member, 1998-2005.
* Research in Education and the Behavioral Sciences, chair, 2005-2007
* Conference 2007 Program Planning, co-chair, 2005-2007.
* Member-at-Large (elected), 2003-2005, held 2004 mid-winter program, "Update on the ERIC reauthorization" and planned and facilitated 2005 midwinter Current Topics Discussion program.
* Nominating Committee, member, 2003-2005, juried vita and put forward candidates for section slate.
* Conference 2004 Program Planning Committee, member, 2002-2004, coordinated "Pedagogy of the Online Learner" program.
* Conference 2002 Program Planning Committee, co-chair, 2000-2002, "Games Academics Play: Mastering the Social Psychology of

Librarian/Faculty Relationships."

* Instruction for Educators Committee, member, 2001-2003, contributed to article for committee.
Curriculum Materials Committee, member, 1997-1998, 1999-2001.
Completed survey of accrediting agencies for standards for curriculum materials centers, 1997-1998.
Created Web site for electronic journals designed for educators and curriculum materials centers, 1996-2001.
Curriculum Materials Committee, chair, 1998-1999.
Ran meetings, expanded committee Web site, presented slate of members for ad hoc committee to update the directory, 1998.
Implemented ad hoc committee to create standards for the management of curriculum materials centers, 1999.
Curriculum Materials Directory Committee, member, 1999-2001, resulted in publication of 5th edition.
EBSS Current Topics Discussion Forum, facilitator, 1999.

New Members Round Table
Local Arrangements Committee, member, 1997-1998, planned and presented at conference orientation.

Machine Assisted Reference Services
Management Committee, member, 1995-1996, program planning and committee consolidation.

Library Information Technology Association
Staffed information booth at Annual Conference, 1994.

Southeastern Library Association

* President's Committee, 2003-2005.
Outstanding Southeastern Author Award Committee, nominated and juried submissions for award, 1998-2002.
Created official Web site for 1996 Southeastern Library Association Biennial Conference for Local Arrangements Committee, with Steven Harris, 1995-1996.

Louisiana Library Association
Academic Section
Legislative Committee, hosted gubernatorial candidates, 1996.
Ballot Committee, certified and reported election results, 1994, 1996.

Other Professional Service

* American Educational Research Association
Graduate Student Liaison to UCF College of Education, 2002-present.

* Reviewer for promotion and tenure, University of Nevada – Las Vegas library faculty member (Jennifer Fabbi), 2004.

* Presented and/or co-hosted professional development sessions for junior faculty, library students' information session, and negotiating the ALA conference sessions for library staff, 2003-2004.

EDUCOM Conference, volunteer, 1998.

Other Service

* Educational Testing Service's Information and Communication Technology test, field test administrator, November 2004.
* Downtown Orlando Partnership, A Gift for Teaching fundraiser, volunteer, 2001.

Relevant Memberships

United Faculty of Florida, UCF Chapter, 1998-2001, 2002-2005.

PROFESSIONAL DEVELOPMENT

Summary of Attendance at Professional Meetings

* American Educational Research Association, Annual Conference, 2002-2003, 2005.
* American Library Association, Annual Conference, 1991, 1994, 1996-2005.
* American Library Association, Midwinter Conference, 1996-2005.
  Charleston Conference (collection development), 1998-2000.
  Consortium for Social Responsibility and Character in Education Conference, 2000.
* Florida Educational Technology Conference, 2005.
  Florida Library Association, Annual Conference, 1999-2001.
  Louisiana Library Association, Annual Conference, 1994-1998.
  Southeastern Library Association, Biennial Conference, 1994, 1996, 2000, 2002.
  Special Libraries Association, Annual Conference, 1993.

Workshops and Training Attended

American Educational Research Association
* Professional development workshop held during annual conference.
  "Making your data a resource for other researchers," April 12, 2005.

  American Library Association/Association of Colleges and Research Libraries
  "Information literacy across the curriculum: Using the information literacy standards as a blueprint for strategic planning," June 2005.

Other Professional Development

* Professional development leave granted for educational and research purposes, leave granted as half-time status for academic year 2001-2002 by UCF Office of Academic Affairs, Summer 2001.

* UCF Faculty Summer Institute 2001, stipend awarded in the amount of $1,000 by the Faculty Center for Teaching and Learning for development of a Web-based tutorial for education students, Summer 2001.

LIST OF REFERENCES

Ackerson, L. G., Howard, J. G., & Young V. E. (1991). Assessing the relationship between library instruction methods and the quality of undergraduate research. *Research Strategies, 9*(2), 139-141.

American Association for Higher Education [AAHE]. (2005). *9 principles of good practice for assessing student learning*. Retrieved January 10, 2005 from http://www.aahe.org/assessment/principl.htm

American Association of School Librarians [AASL] & Association for Educational Communications and Technology [AECT]. (1998). *Information literacy standards for student learning.* Chicago, IL: American Library Association.

American Psychological Association [APA]. (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

Anderson, T. H. (1974). Cloze measures as indices of achievement comprehension when learning from extended prose. *Journal of Educational Measurement, 11*, 83-92.

Associated Colleges of the South [ACS]. (2003). *Information fluency working definition*. Retrieved February 24, 2005, from: http://www.colleges.org/~if/if_definition.html

Association of American Colleges and Universities [AACU]. (2005). *Liberal education outcomes: A preliminary report on student achievement in college.* Washington, DC: Author.

Association of College and Research Libraries [ACRL], Task Force on Information Literacy Competency Standards in Higher Education. (2000). *Information literacy competency standards for higher education*. Retrieved June 8, 2004, from http://www.ala.org/acrl/ilcomstan.html

Bandura, A. J. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-215.

Barclay D. (1993). Evaluating library instruction: Doing the best you can with what you have. *RQ, 33*, 195-202.

Beile, P. M. & Boote, D. N. (2002). Library instruction and graduate professional development: Exploring the effect of learning environments on self-efficacy and learning outcomes. *Alberta Journal of Educational Research, 68*(4), 364-367.

Beile, P. M., Boote, D. N., & Killingsworth, E. K. (2003). *Characteristics of education doctoral dissertation references: An inter-institutional analysis of review of literature citations.* Paper presented at the American Educational Research Association Annual Conference, Chicago, IL. (ERIC Document Reproduction Service No. ED478598).

Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research, 56*(1), 137-172.

Bloom, B. S., Madaus, G. F., and Hastings, J. T. (1981). *Evaluation to improve learning.* New York: McGraw-Hill.

Board, C. & Whitney, D. R. (1972). The effect of selected poor item-writing practices on test difficulty, reliability and validity. *Journal of Educational Measurement, 9,* 1972, 225-233.

Bober, C., Poulin, S., & Vileno, L. (1995). Evaluating library instruction in academic libraries: A critical review of the literature, 1980-1993. In L. M. Martin (Ed.), *Library instruction revisited: Bibliographic instruction comes of age* (pp.53-71). New York: The Haworth Press.

Boote, D. N. & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher, 34*(6), 3-15.

Bradley, J. (1993). Methodological issues and practices in qualitative research. *Library Quarterly, 63*(4), 431-449.

Breivik, P. S. (1985). A vision in the making: Putting libraries back in the Information Society. *American Libraries, 16*(1), 723.

Breivik, P. S. (1991). Literacy in an Information Society. *Information Reports and Bibliographies, 20*(3), 10-14.

Bren, B., Hillemann, B., & Topp, V. (1998). Effectiveness of hands-on instruction of electronic resources. *Research Strategies, 16*(1), 41-51.

Brown, J. D. (2000). What is construct validity? *JALT Testing & Evaluation SIG Newsletter*, *4*(2), 7-10. Retrieved March 23, 2005, from http://www.jalt.org/test/bro_8.htm

Budd, J. M. (1995). An epistemological foundation for library and information science. *Library Quarterly, 65*(3), 295-318.

Cameron, L. (2004). Assessing information literacy. In I. F. Rockman (Ed.), *Integrating information literacy into the higher education curriculum: Practical models for transformation* (pp. 207-236). San Francisco: Jossey Bass.

Chadley, O. & Gavryck, J. (1989). Bibliographic instruction trends in research libraries. *Research Strategies, 7*, 106-113.

Clark, L. A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319.

Cohen, J. & Cohen, P. (1983). *Multiple regression/correlation in the behavioral sciences*. 2nd Ed. Hillsdale NY: Erlbaum and Associates.

Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage Publications.

Daugherty, T. K. & Carter, E. W. (1997). Assessment of outcome-focused library instruction in Psychology. *Journal of Instructional Psychology, 24*(1), 29-33.

Davis, F. B., & Diamond, J. J. (1974). The preparation of criterion-referenced tests. In C. W. Harris, M. C. Alkin, & W. J. Popham (Eds.). *Problems in criterion-referenced measurement* [Center for the Study of Evaluation monograph series, No.3] (pp. 116-138). Los Angeles: Center for the Study of Evaluation.

Dillman, D. A. (1999). *Mail and telephone surveys: The tailored design method*. New York: John Wiley and Sons.

Dillman, D. A., Tortora, R. D., Conradt, J., & Bowker, D. (1998). *Influence of plain vs. fancy design on response rates for web surveys*. Retrieved February 22, 2005, from Washington State University, Social and Economic Sciences Research Center Web site: http://survey.sesrc.wsu.edu/dillman/papers/asa98ppr.pdf

Eadie, T. (1992). Beyond immodesty: Questioning the benefits of BI. *Research Strategies, 10*, 105-110.

Ebel, R. L. (1968). The value of internal consistency in classroom examinations. *Journal of Educational Measurement, 5*, 71-73.

Ebel, R. L. (1972). Why is a longer test usually a more reliable test? *Educational and Psychological Measurement, 32*, 24-253.

Eco, U. (1983). *The name of the rose* (W. Weaver, Trans.). San Diego, CA: Harcourt, Brace, Jovanovich. (Original work published 1980).

Education and Behavioral Sciences Section [EBSS]. (1992). Information retrieval and evaluation skills for education students. *C&RL News*, *9*, 583-588.

Educational Testing Service [ETS]. (2004). *ICT literacy*. Retrieved February 25, 2005, from: http://www.ets.org/ictliteracy/

Edwards, S. (1994). Bibliographic instruction research: An analysis of the journal literature from 1977 to 1991. *Research Strategies, 12*, 68-78.

Electronic Publishing Initiative at Columbia [EPIC]. (2004). *Online survey of college students: Executive summary*. Retrieved June 22, 2004, from Columbia University, Electronic Publishing Initiative at Columbia Web site: http://www.epic.columbia.edu/eval/find09/find09.html

Elliot, S. N., Kratochwill, T. R., Littlefield, J. and Travers, J. F. (1996). *Educational psychology: Effective teaching, effective learning* (2nd Ed.). Madison, WI: Brown and Benchmark.

Erlendsson, J. (2003). *Essays on Derek John de Solla Price*. Retrieved February 23, 2005, from: http://www.hi.is/~joner/eaps/sollapri1.htm

Fox, L. M. & Weston, L. (1993). Course-integrated instruction for nursing students: How effective?. *Research Strategies, 11*, 89-99.

Franklin, G. & Toifel, R. C. (1994). The effects of BI on library knowledge and skills among Education students. *Research Strategies, 12*, 224-237.

Gardner, P. L. (1970). Test length and the standard error of measurement. *Journal of Educational Measurement, 7*, 271-273.

Grassian, E. S. & Kaplowitz, J. R. (2001). *Information literacy instruction: Theory and practice*. New York: Neal-Schuman Publishers.

Gratch Lindauer, B. & Brown, A. (2004). Developing a tool to assess community college students. In I. F. Rockman (Ed.), *Integrating information literacy into the higher education curriculum: Practical models for transformation* (pp. 165-206). San Francisco: Jossey Bass.

Greer, A., Weston, L. & Alm, M. L. (1991). Assessment of learning outcomes: A measure of progress in library literacy. *College & Research Libraries, 52*, 549-557.

Groves, R. M. (1989). *Survey errors and survey costs.* New York: John Wiley.

Hagner, P. A. & Hartman, J. L. (2004). *Faculty engagement, support and scalability issues in online learning.* Paper presented at the Academic Impressions Web Conference, January 14, 2004. [Retrieved from compact disk video of conference.]

Hakstian, A. R. & Kansup, W. (1975). A comparison of several methods of assessing partial knowledge in multiple-choice tests: II. Testing procedures. *Journal of Educational Measurement, 12*, 231-239.

Harris, C. W. (1974a). Problems of objectives-based measurement. In C. W. Harris, M. C. Alkin, & W. J. Popham (Eds.). *Problems in criterion-referenced measurement* [Center for the Study of Evaluation monograph series, No.3] (pp. 83-94). Los Angeles: Center for the Study of Evaluation.

Harris, C. W. (1974b). Some technical characteristics of mastery tests. In C. W. Harris, M. C. Alkin, & W. J. Popham (Eds.). *Problems in criterion-referenced measurement* [Center for the Study of Evaluation monograph series, No.3] (pp 98-115). Los Angeles: Center for the Study of Evaluation.

Hatcher, L. (1994). *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. Cary, NC: SAS Institute.

Hernon, P. & Dugan, R. E. (2004). *Outcomes assessment in higher education: Views and perspectives.* Westport, CT: Libraries Unlimited.

Herring, M. Y. (2001). 10 reasons why the Internet is no substitute for the library. *American Libraries*, *32*(4), 76-78.

Horn, J. L. (1966). Some characteristics of classroom examinations. *Journal of Educational Measurement, 3,* 293-295.

Horn, J. L. (1968). Is it reasonable for assessments to have different psychometric properties than predictors? *Journal of Educational Measurement, 5,* 75-77.

Ingwersen, P. (1982). Search procedures in the library – analysed from the cognitive point of view. *Journal of Documentation, 38*, 165-191.

International Society for Technology in Education [ISTE], National Educational Technology Standards for Teachers [NETS*T]. (2000). *Educational technology standards and performance indicators for all teachers*. Retrieved July 19, 2004, from: http://cnets.iste.org/teachers/t_stands.html

Interstate New Teacher Assessment and Support Consortium [INTASC]. (1992). *Model standards for beginning teacher licensing, assessment and development: A resource for state dialogue*. Retrieved May 22, 2004, from: http://www.ccsso.org/content/pdfs/corestrd.pdf

James Madison University. (2004, June). *Information literacy test.* Retrieved July 15, 2005, from James Madison University, Center for Assessment and Research Studies Web site: http://www.jmu.edu/assessment/wm_library/ILT.pdf

Katz, I. (2005, June). *The ICT Literacy Assessment test: An update*. Report presented at the annual conference of the American Library Association, Chicago, IL.

Kehoe, J. (1995). Basic item analysis for multiple-choice tests. *Practical Assessment, Research & Evaluation*, *4*(10). Retrieved September 20, 2004, from http://pareonline.net/getvn.asp?v=4&n=10

Kerlinger, F. N. & Lee, H. B. (2000). *Foundations of behavioral research* (4th Ed.). Fort Worth, TX: Harcourt College Publishers.

Kennedy, M. M. (1997). The connection between research and practice. *Educational Researcher, 26*(7), 4-12.

Kohl, D. F. & Wilson, L. A. (1986). Effectiveness of course-integrated bibliographic instruction in improving coursework. *RQ, 26*, 206-211.

Kranich, N. C. (2000). Building partnerships for 21st-century literacy. *American Libraries*, 31(8), 7.

Kuhlthau, C. C. (1993). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.

Kunkel, L. R., Weaver, S. M., & Cook, K.N. (1996). What do they know?: An assessment of undergraduate library skills. *Journal of Academic Librarianship, 22*, 430-434.

Kuyper, L. A. & Dziuban, C. D. (1984). Passing scores and the Medical Record Administration Registration Examination. *Journal of AMRA, 55*, 29-30.

Linacre, J. M. (2004). Test validity and Rasch measurement: Construct, content, etc. Retrieved, March 23, 2005, from http://www.rasch.org/rmt/rmt181h.htm

Linn, R. L. and Gronlund, N. (1995). *Measurement and assessment in teaching* (7th Ed.). New York: Macmillan.

Lord, F. M. (1974). Quick estimates of the relative efficiency of two tests as a function of ability level. *Journal of Educational Measurement, 11*, 247-254.

Lubans, Jr., J. (1983). *Educating the public library user.* Chicago: American Library Association.

Mager, R. F. (1997). *Preparing instructional objectives: A critical tool in the development of effective instruction* (3rd Ed.). Atlanta: Center for Effective Performance.

Maki, P. I. (2002). Developing an assessment plan to learn about student learning. *Journal of Academic Librarianship, 28*(1/2), 8-13.

Martin, B.L. (1989). A checklist for designing instruction in the affective domain. *Educational Technology, 29*, 7-15.

Maughan, P. D. (2001). Assessing information literacy among undergraduates: A discussion of the literature and the University of California-Berkeley experience. *College & Research Libraries, 62,* 71-85.

Mensching, T. B. (1987). Reducing library anxiety and defining teaching. *Research Strategies, 5*, 146-148.

Middle States Commission on Higher Education [MSCHE]. (2002). *Characteristics of excellence in higher education: Eligibility requirements and standards for accreditation.* Philadelphia, PA: Middle States Commission on Higher Education.

Morner, C. J. (1993). A test of library research skills for education doctoral students. (Doctoral dissertation, Boston College, 1993). *Dissertation Abstracts International A, 54/*6, 2070.

Nahl-Jakobovits, D.& Jakobovits, L. A. (1993). Bibliographic instructional design for information literacy: Integrating affective and cognitive strategies. *Research Strategies, 11,* 73-88.

National Council for Accreditation of Teacher Education [NCATE]. (2002). *Professional standards for accreditation of schools, colleges, and departments of education*. Retrieved June 3, 2005, from http://www.ncate.org/documents/unit_stnds_2002.pdf

National Research Council, Committee on Information Technology Literacy. (1999). *Being fluent with information technology.* Washington, DC: National Academy Press.

New England Association of Schools and Colleges [NEASC]. (2005). *Standards for accreditation*. Retrieved August 5, 2005, from http://www.neasc.org/cihe/standards_for_accreditation_2005.pdf

Nitko, A. J. (1970). Criterion-referenced testing in the context of instruction. In *Testing in turmoil: A conference on problems and issues in educational measurement*. Greenwich, CT: Educational Records Bureau.

Nitko, A. J. (1974). Problems in the development of criterion-referenced tests: The IPI Pittsburgh experience. In C. W. Harris, M. C. Alkin, & W. J. Popham (Eds.). *Problems in criterion-referenced measurement* [Center for the Study of Evaluation monograph series, No.3] (pp. 59-82). Los Angeles: Center for the Study of Evaluation.

Northwest Commission on Colleges and Universities [NWCCU]. (2003). *Accreditation standards*. Retrieved July 15, 2005, from http://www.nwccu.org/Standards%20and%20Policies/Accreditation%20Standards/Accreditation%20Standards.htm

Novick, M. R. & Lewis, C. (1974). Prescribing test length for criterion-referenced measurement. In C. W. Harris, M. C. Alkin, & W. J. Popham (Eds.). *Problems in criterion-referenced measurement* [Center for the Study of Evaluation monograph series, No.3] (pp. 139-158). Los Angeles: Center for the Study of Evaluation.

O'Connor, L. G., Radcliff, C. J., & Gedeon, J. (2002). Applying systems design and item response theory to the problem of measuring information literacy skills. *College and Research Libraries, 63*(6), 528-543.

Patterson, C. D., & Howell, D. W. (1990). Library user education: Assessing the attitudes of those who teach. *RQ, 29*, 513-523.

Pew Internet & American Life Project. (2005). *Search engine users: Internet searchers are confident, satisfied, and trusting – but they are also unaware and naïve*. Retrieved June 27, 2005, from: http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf

Popham, W. J. (1974). Selecting objectives and generating test items for objectives-based tests. In C. W. Harris, M. C. Alkin, & W. J. Popham (Eds.). *Problems in criterion-referenced measurement* [Center for the Study of Evaluation monograph series, No.3] (pp. 13-25). Los Angeles: Center for the Study of Evaluation.

Popham, W. J. & Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement, 6*, 1-9.

Postman, N. (2004). The Information Age: A blessing or a curse? *Harvard International Journal of Press/Politics*, 9(2), 3-10. (Reprinted from the Joan Shorenstein Center on the Press, Politics and Public Policy, 1995).

Project SAILS (2001) *Project SAILS: Project for the standardized assessment of information literacy skills*. Retrieved March 5, 2004, from Kent State University, Project SAILS Web site: http://sails.lms.kent.edu/index.php

Rabine, J. L. & Cardwell, C. (2000). Start making sense: Practical approaches to outcomes assessment for libraries. *Research Strategies, 17*(4), 319-335.

Radcliff, C. J. (2005, June). *Project SAILS update*. Report presented at the annual conference of the American Library Association, Chicago, IL.

Rader, Hannelore. (1991). Information literacy: A revolution in the library. *RQ, 31,* 25-29.

Rader, Hannelore. (2000). A silver anniversary: 25 years of reviewing the literature related to user instruction. *Reference Services Review, 28*(3), 290-296.

Reitz, J. M. (2004). *Dictionary for library and information science*. Westport, CT: Libraries Unlimited.

Ren, W. H. (2000). Library instruction and college student self-efficacy in electronic information searching. *Journal of Academic Librarianship, 26*, 323-328.

Rettig, J. & Hagen, S. K. (2003). Stakeholders and strategies in information fluency. *Transformations: Liberal Arts in the Digital Age, 1*(1). Retrieved February 24, 2005, from: http://www.colleges.org/transformations/index.php?q=node/view/7

Rockman, I. F. (2004). Introduction: The importance of information literacy. In I. F. Rockman (Ed.), *Integrating information literacy into the higher education curriculum: Practical models for transformation* (pp. 1-28). San Francisco: Jossey Bass.

Russell, M. & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives, 5*(3). Retrieved March 15, 2005, from: http://epaa.asu.edu/epaa\v5n3/

Salony, M. F. (1995). The history of bibliographic instruction: Changing trends from books to the electronic world. *The Reference Librarian, 51/52*, 31-51.

Schuck, B. R. (1992). Assessing a library instruction program. *Research Strategies, 10*, 152-160.

Shapiro, J. J. & Hughes, S. K. (1996). Information as a liberal art. *Educom Review, 31*(2), 31-36.

Shulman, L. S. (1999). Professing educational scholarship. In E. C. Lagemann & L. S. Shulman (Eds.). *Issues in education research: Problems and possibilities* (pp. 159-164). San Francisco: Jossey Bass.

Simon, G. B. (1969). Comments on "Implications of criterion-referenced measurement." *Journal of Educational Measurement, 6*, 259-260.

Skager, R. W. (1974). Creating criterion-referenced tests from objectives-based assessment systems: Unsolved problems in test development, assembly, and interpretation. In C. W. Harris, M. C. Alkin, & W. J. Popham (Eds.). *Problems in criterion-referenced measurement* [Center for the Study of Evaluation monograph series, No.3] (pp. 47-58). Los Angeles: Center for the Study of Evaluation.

Sutton, B. (1993). The rationale for qualitative research: A review of principles and theoretical foundations. *Library Quarterly, 63*(4), 411-430.

Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion-referenced tests: A decision-theoretic formulation. *Journal of Educational Measurement, 11*, 263-267.

Thompson, B. & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*(2), 197-208.

Tierno, M. J. & Lee, J. H. (1983). Developing and evaluating library research skills in education: A model for course-integrated bibliographic instruction. *RQ, 22*, 284-291.

Tunon, J. (1999). *Integrating bibliographic instruction for distance education doctoral students into the Child and Youth Studies program at Nova Southeastern University.* Unpublished doctoral practicum, Nova Southeastern University, Fort Lauderdale, FL. (ERIC Document Reproduction Service No. ED440639).

University of Central Florida [UCF], Office of Institutional Research. (2004). *University of Central Florida Headcount by Major and Level: Final fall 2004 (all students).* Retrieved May 22, 2005, from http://www.iroffice.ucf.edu/enrollment/2004-05/1ffal04_allstudents_w.pdf

Varian, H. (2003). *How much information? 2003: Executive summary*. Retrieved February 23, 2005, from University of California, Berkeley, School of Information Management and Systems Web site: http://www.sims.berkeley.edu/research/projects/how-much-info-2003

Varma, S. (n.d.). *Computing and using point biserials for item analysis.* Morgan Hill, CA: Educational Data Systems.

Wilbur, P. H. (1970). Positional response set among high school students on multiple-choice tests. *Journal of Educational Measurement, 7*, 161-163.

Zaporozhetz, L. E. (1987). The dissertation literature review: How faculty advisors prepare their doctoral candidates. (Doctoral dissertation, University of Oregon, 1987). *Dissertation Abstracts International A, 48/11,* 2820.

Zurowski, P. G. (1974). *The information service environment: Relationships and priorities.* Washington, DC: National Commission on Libraries and Information Science.