Running head: DEVELOPMENT AND VALIDATION OF THE ILAS-ED

Development and Validation of the Information Literacy Assessment Scale for Education (ILAS-ED)

Penny Beile

University of Central Florida

AERA Annual Conference

Montreal, Canada

April 12, 2005

Abstract

No population exists where it is more important to produce information literate individuals than teacher candidates, yet few would suggest that practitioners newly entering the field are adequately prepared to model and teach information literacy to their students. Consequently, information literacy has recently been established as a key outcome by a number of teacher education accrediting bodies and professional associations. Only in the last few years has there been an attempt to develop a standardized scale to assess general information literacy skills, and at the time of this writing no standardized tool exists that measures the information literacy levels of teacher candidates.

This study documents the development and validation of a standardized instrument to measure teacher candidates' information literacy skills levels based on the International Society for Technology in Education's 2000 *National Educational Technology Standards for Teachers* and the Association of College and Research Libraries' 2000 *Information Literacy Competency Standards for Higher Education*. Undergraduate students enrolled in the teacher education program at the University of Central Florida were identified and asked to complete a test consisting of 22 multiple-choice test items and 13 demographic and self-percept items. A number of procedures designed to enhance validity and reliability of the scale were integrated throughout its development. Results of the test were also submitted to analysis.

This project is part of a national initiative to develop standardized information literacy assessment tools specific to a discipline, and is spear-headed by the Project for the Standardized Assessment of Information Literacy Skills and the Institute for Library and Information Literacy Education. Use of the instrument described herein will allow librarians and teaching faculty a means to inform curricular and instructional decisions, and results can be used for internal and external benchmarking of education students' information literacy skills levels.

Development and Validation of the Information Literacy Assessment Scale for Education (ILAS-ED)

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. Mary Kennedy (1997) argues that few teachers use the scholarly literature to inform their professional practice as they do not perceive the connection between research and practice. Kennedy proposes that initiatives such as ERIC have been successful in facilitating physical access to the literature, but concedes that conceptual barriers still exist. Other researchers have likewise reported cognitive or conceptual discrepancy regarding scholarly information access and use. Reported research suggests 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, 2003), or feel insufficiently prepared to successfully negotiate the information environment to locate, evaluate, and cite needed sources (Zaporozhetz, 1987).

Information literacy, with its emphasis on critical thinking and problem-solving skills as they relate to an individual's information need, has recently been recognized by educators and business professionals alike as fundamental to success in a rapidly changing, technology and information intensive environment. Although no population exists where it is more important to produce information literate individuals than teacher candidates, few would suggest that practitioners newly entering the profession are adequately prepared to model and teach information literacy to their students. Perhaps this is one reason why the National Council for the Accreditation of Teacher Education (NCATE, 2002), American Association of School Librarians and Association for Educational Communications and Technology (AASL / AECT, 1998), and International Society for Technology in Education (ISTE, 2000), have recently adopted information literacy as a key outcome for teacher education students. Additionally, these standards generally recommend that information literacy instruction be viewed as a cumulative and continuous process that is woven throughout the curriculum (Grassian & Kaplowitz, 2001; Hagner & Hartman, 2004; ISTE, 2000; Middle States Commission on Higher Education, 2002), therefore implying that the integration of information literacy instruction is the responsibility of all academicians.

Concurrent to these developments, the Association of College and Research Libraries (ACRL, 2002) developed and approved information literacy competency standards for higher education. These standards have served to unite disparate instructional initiatives of various academic libraries and associations, and also have clarified the library's role in supporting institutional information literacy instruction efforts. Although accreditation standards assign responsibility for information literacy instruction to all academic faculty, the library's ability to customize information literacy instruction to individual programmatic needs places the library central to delivery of information literacy instruction in the academy. In addition to the obvious value for curriculum and instruction planning, the ACRL standards offer possibility for unified assessment efforts around the country (O'Connor, Radcliff, & Gedeon, 2002).

Assessment data can provide meaningful information for both internal and external benchmarking and in this era of accountability and shrinking resources institutions are challenged to provide evidence that their instructional programs positively impact student learning. At the local level, assessment can help determine if teacher candidates possess adequate information literacy skills and knowledge, in turn contributing to the evaluation and revision of institutional information literacy instruction programs. Perhaps even more vital, assessment results offer another data point regarding institutional performance for accreditation reviews.

Purpose of the Study

A number of researchers have developed tools for measuring students' cognitive or affective changes after library instruction, yet the majority of these instruments have been developed for local use only and have not been submitted to rigorous scrutiny. No standardized assessment instrument exists at the time of this writing that measures information literacy levels of teacher candidates. 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" (p. 6), the ACRL (2000) information literacy task force explicitly called for the development of assessment instruments that are unique to the academic discipline.

The purpose of this study was to extend evaluation of information literacy learning in teacher education programs by developing and validating an objective assessment instrument that meets the three-fold challenge of measuring teacher candidates' cognitive knowledge of information literacy, promising potential to be used across differing institutional settings, and providing an instrument specific to the discipline of education.

Review of Literature

Evaluation of library information literacy programs is frequently discussed in the literature, yet rigorous assessment studies are not often reported (Bober, Poulin, & Vileno, 1995). Thomas Eadie (1992) suggests that evaluation studies tend to report on student perceptions or "user satisfaction" of library instruction and/or resources rather than learning outcomes. In a content analysis study of library instruction related articles, Edwards (1994) describes a three-fold increase in the number of articles published from 1977 through 1991, yet an annual review of library instruction research performed by Hannelore Rader (2000) reveals that a considerable number of publications tend to be program descriptions. Surveys from the 1970s and 1980s confirm that evaluation was not a major component of library and information literacy instruction (Bober et al., 1995; Chadley & Gavryck, 1989). These conditions led Werking (1980) to conclude that systematic, formal evaluation has not occurred to any significant degree, and the ensuing 25 years have revealed very little improvement.

Research indicates several barriers to formal instruction evaluation. Patterson and Howell (1990) state that most library schools do not offer classes on instructional assessment, thus leaving many librarians feeling they are ill-prepared to properly conduct assessment studies. Others see formal evaluation as too complex or too time consuming,

and may cite the lack of institutional support (Eadie, 1992). Eadie adds that often evaluation is perceived as one more responsibility on an already excessive workload. In addition, library information literacy instructors may be unwilling to include assessment in their sessions because it reduces the amount of material that can be included in the limited class time available to them (Grassian & Kaplowitz, 2001).

Despite the criticism levied toward the state of instructional assessment, several studies are described in the literature. These studies, however, differ on what was being assessed, the methodology used for assessment, and the inferences reached based on analysis of the data. Reported studies tend to fall into two categories; those that investigate instructional impact on the affective domain, and those that focus on cognitive outcomes. Researchers who report positive post-instruction statistical significance, whether for affective or cognitive impact, include Leighton and Markham (1991), Tierno and Lee (1983), Daugherty and Carter (1997), Franklin and Toifel (1994), Dykeman and King (as cited in Bober et al., 1995), Schuck (1992), and Ren (2000). Other research (c.f.e, Fox & Weston, 1993; Maughan, 2001; and Greer, Weston, & Alm, 1991) has failed to find a statistically significant relationship between instruction and attitudinal or learning gains.

In every instance these studies reported using a locally-produced evaluation tool that had not been submitted to validity and reliability analysis. Acknowledging this limitation, Barclay (1993) adds that there are no widely accepted standardized tests for evaluating library use at the college level. Bober et al. (1995) also concede this is true, and caution that use of locally produced tests may increase unreliability or bias. When discussing impediments to formal information literacy instruction evaluation, none are as problematic as the lack of a global assessment instrument.

Attempting to meet this need is the purpose of Project SAILS (2001), a federally funded initiative devoted to developing an information literacy assessment instrument that has been proven valid and reliable, is easy to administer, is standardized, allows for use at any institution, and provides for both internal and external benchmarking. To date, Project SAILS has developed a test bank of approximately 150 general information literacy test items, and at last count had 77 institutions participating in assessment of their information literacy instruction programs. Project SAILS test items are designed to evaluate information literacy skills that are appropriate to an undergraduate learner; these skills are general in that they are not specific to any particular discipline.

However, based on ACRL's (2000) appeal for the development of assessment instruments and strategies unique to the academic discipline, Project SAILS forwarded a call for participation to develop discipline-specific modules of the SAILS instrument (Project SAILS, 2001). A project team from the University of Central Florida responded to the announcement and was awarded a fellowship to develop education-specific test items to populate the Project SAILS test bank. After completing all requirements of the fellowship, the items provided the foundation for developing a more parsimonious and more easily administered assessment tool. It is the development and validation of this instrument that is described in this paper.

Methodology

Methods are comprised of two sections, and are identified as Phase I and Phase II. Phase I describes work performed for the Project SAILS fellowship, which was awarded to develop test content and populate a test item bank for education-related information sources, while Phase II explains procedures for the subsequent development of an assessment scale. Objective, outcomes-based assessment measures require a number of procedures to attest to their credibility, among them are checks for validity and reliability. The following methods detail efforts to enhance validity and reliability of the scale.

Phase I – Project SAILS-supported

Test content. The most comprehensive standards, in that a number of learning outcomes and objectives have been developed to accompany them, are the ACRL *Information Literacy Competency Standards for Higher Education* (2000). As such, these standards were chosen as the basis for the study. However, standards that apply to teacher education accreditation efforts also exist. Therefore, themes pertinent to information literacy that run throughout the ISTE *National Educational Technology Standards for Teachers* (*NETS*T*), which are relied upon by NCATE, were aligned with ACRL standards and objectives to form a basis for test content development. The four broad areas of information competence suggested by the *NETS*T* include identifying, evaluating, and selecting finding tools; demonstrating knowledge of general search strategies; evaluating and selecting sources; and demonstrating knowledge of legal and ethical practices.

Test item construction. After Project SAILS review and approval of the test content parameters, writing of items designed to measure students' levels of information literacy skills as they relate to identified objectives commenced. Project SAILS personnel suggested that development teams identify 30 to 40 objectives, and then write items to assess cognitive knowledge of those objectives. Preliminary item writing resulted in a bank of 58 test items. Project team review revealed the need for additional items in the area of ethical use of information. Four more test items were written to address this category, bringing the total to 62 items.

Further test development. Individual testing was conducted with a combination of six newly hired and continuing library student assistants. The students 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. The think-aloud protocol served to identify language and conceptual constraints and helped to clarify items.

Content validity. To enhance content validity, a panel of five experts in the field reviewed items and rated them on a scale of 0 (absent the quality) to 4 (fully expresses the quality) for content accuracy, or alignment to ACRL objectives, clarity of item, and institutional objectivity. Items with a mean average score of 2.0 or below in any category were reviewed by the researchers, and either rewritten or marked as potentially

problematic for Project SAILS personnel. Upon completion of one-on-one testing and receipt of content reviewers' comments, the project team worked on revising items and formatting them for survey.

Pilot testing. Students enrolled in two education classes were asked to complete the pilot test, which resulted in 29 usable surveys. Student responses were entered into a spreadsheet and item analysis procedures, including difficulty level, item discrimination index, and distractor analysis, were performed. Problematic items were flagged and 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. Researchers retained rights to further use of the items and publication relating to them.

Phase II – Instrument Development

Item reduction. Sampling across all the content clusters, and seeking a range of difficulty levels, the 62 content items initially developed for the Project SAILS test bank were reduced to 22 items. The multiple-choice format was retained as it lends itself to being answered and scored more quickly than constructed response items. Demographic and other non-content area items were added to the test. Differences in scores may be affected by any number of factors represented by the demographic and non-content area items, and data from these questions were used as the basis of independent variables analysis described in the Results section.

Population and Setting. Students who participated in testing were from University of Central Florida, a public, metropolitan university with enrollments of over 40,000 students. As of September 2004, 3,053 undergraduate students (83.85% female and 16.15% male) were enrolled in the College of Education (UCF Office of Institutional Research, 2004). An email invitation to complete a web-based test was extended to the whole population. No incentives were offered to web participants, possibly explaining the rather low response rate of 3%, or 92 usable surveys which were received. Subsequent efforts to bolster the sample size included placing signage in a busy lobby of an education building and offering a $5 incentive to complete the test. This netted 80 more surveys, for a total of 172 responses.

Test administration. 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 web version of the test contained the same items as the print version, but administration differed slightly. For the web-administered test, 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 online. Students who completed the print test received a test packet containing the test, test directions, a scantron, and two informed consent forms. Students completed the test in the library. The length of time it took to complete the test ranged from 20 to 25 minutes, on average.

Statistics that describe the sample. Descriptive statistics that show the frequencies of various independent variables for the 172 participants who completed the ILAS-ED at the University of Central Florida follow. Missing values are indicated, and in all other cases N=172. Table 1 gives the breakdown by gender, ethnicity, student classification, and length of enrollment at UCF.

Participants were 34 males (19.77%) and 136 females (79.07%), two students did not indicate their gender. The majority of the participants who responded to the question of ethnicity were White or European American (81.39%), followed by Black or African-American (8.14%), Hispanic or Latino (7.56%), and Asian or Asian American (2.3%). Only one person listed "other," and that person indicated they were Arab. Statistics that describe the sample are further reported in *Analysis of Test Data by Respondents' Characteristics*, which is located in the Results section. The sample is fairly representative of the undergraduate teacher education majors at the University of Central Florida.

Table 1. *Frequencies of the Sample*

	Category	Number	Percent
Gender	Male	34	19.77
	Female	136	79.07
	Missing	2	1.16
	Total	172	100.00
Ethni-city	White or European American	140	81.39
	Black or African American	14	8.14
	Hispanic or Latino	13	7.56
	Asian or Asian American	4	2.32
	Other (Arab)	1	.59
	Total	172	100.00
Student Classi-fication	Freshman	12	7.0
	Sophomore	10	5.8
	Junior	48	27.9
	Senior	80	46.5
	Missing	22	12.8
	Total	172	100.00
UCF Enroll-ment	Less than one year	41	23.84
	1 or 2 years	55	31.98
	2 or 3 years	50	29.07
	4 or more years	23	13.37
	Missing	3	1.74
	Total	172	100.00

Criterion-related validity. Criterion-related validity, often operationalized as determining accuracy of the measure 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. Ten participants were selected from the pool of candidates who had completed the written portion of the 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% correct, to a high of 19, or 86% correct.

Participants were scheduled for one-half hour time slots at the end of the testing period. The follow-up test was conducted in the curriculum materials library and the in-library test was administered by the researcher. This phase of the testing occurred anywhere from 14 to 20 days after students completed the written test. The in-library test was developed from the written test. Eight test items, representing each of the content clusters, were selected from the original 22 items. Results of the criterion-related validity procedures are described in the following section.

Statistical measures. Descriptive statistics of the test and descriptive statistics of the sample were calculated. Analysis of test items included checking distractors for plausibility and calculating item difficulty levels and discrimination indices. Reliability procedures consisted of stability checks and internal consistency calculation. To measure stability, eleven students were tested twice and results analyzed. Internal consistency was calculated using the Kuder Richardson 20 formula for item-subscale correlations. Factor analysis of the scale and content clusters was conducted.

Results

Upon completion of the Project SAILS fellowship, the researcher sought to develop a briefer test that could be easily administered and scored by classroom or library faculty. Criteria used to reduce the 62 items developed for Project SAILS to the 22 items for the ILAS-ED were based on alignment of the *NETS*T* standards with existing ACRL objectives, rating scores from five content experts, and results of a think aloud protocol and subsequent pilot testing. Thirteen demographic and self-percept questions were added, bringing the final version of the test to 35 questions. The test was administered to 172 students enrolled in an education program at a large urban university and results were submitted to analysis.

Descriptive Statistics for the Test

A comparison of data from the two administration modes reveals more similarities than differences. Table 2 displays descriptive statistics for student test scores from each administration mode and the total sample.

Despite a one point difference in mean scores between students who completed the print-administered test and the web-administered test, there was no statistically significant difference 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. The Kuder Richardson reliability coefficient was also relatively constant. This indicates student scores were fairly consistent across mode of test administration, and that future administrators can be confident in delivering the test as print-based or web-based.

Table 2. *Descriptive Statistics for Two Administration Modes and Total Sample*

	Print Administered	Web Administered	Total Sample
Mean Score	11.44*	12.43*	11.97
Mean Percent	51.99	56.53	54.42
Median	50.00	54.55	54.55
Mode	50.59	50.00	50.00
Range (%)	9-86	14-91	9-91
KR Reliability	.673	.678	.675
St Dev	17.08	16.70	16.98
SEM	1.91	1.74	1.29
Number	80	92	172

*p>.05

The frequency distribution, shown in Table 3 reveals that raw scores ranged from 2 to 20, out of a possible 22. The distribution of scores is fairly normal, with 46% falling into the midrange of 10-14, which closely approximates the second and third quartile. Figure 1 presents a graphical representation of the distribution of scores.

Table 3. *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

Figure 1. *Frequency Distribution of Test Scores*

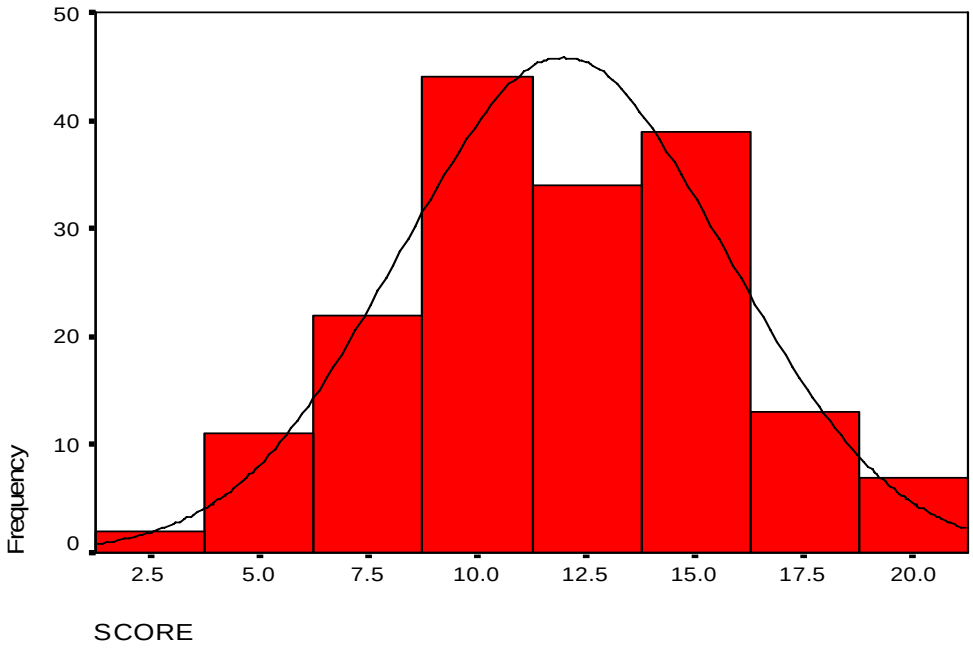

Table 4 shows item level data, including difficulty and discrimination indices and the percent choosing each response for each test stem. In the table, "item number" reflects the actual test item numbering. Items 1 through 6 and 29 through 35 were demographic or self-percept items. "Correct Answer" refers 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, the correct answer and distractors.

A broad range exists in difficulty level for the 22 items. Difficulty levels range from only 32% answering item 6 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 information literacy skills levels. A range of difficulty levels is also dispersed among the four content clusters. The first cluster, identifying, evaluating, and selecting finding tools; contained items with a difficulty range of .32 to .68. The second cluster, demonstrating knowledge of general search strategies; ranged from .39 to .73, while cluster three, evaluating and selecting sources; ranged from .36 to .69 and cluster four, demonstrating knowledge of legal and ethical practices, contained items with a difficulty range of .34 to .89.

The discrimination index 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. As negatively scored items are not

adequately discriminating among knowledge levels, it is generally recommended that they be revised. No negative item discrimination values were uncovered, thus indicating that test items discriminated between high and low scores in the desired direction.

Table 4. *Item Analysis*

				Percent Choosing			
Item Number	Correct Answer	Difficulty	Discrimination	A	B	C	D
7	C	.49	0.317	21	25	49	5
8	D	.32	0.230	34	10	24	32
9	D	.57	0.249	9	8	27	57
10	A	.41	0.271	41	22	35	2
11	D	.39	0.250	37	19	6	39
12	D	.68	0.360	5	7	19	68
13	B	.65	0.166	30	65	4	1
14	A	.60	0.231	60	14	20	6
15	C	.42	0.084	12	21	42	24
16	B	.59	0.445	14	59	10	17
17	C	.73	0.411	10	10	73	6
18	B	.65	0.211	15	65	12	8
19	B	.36	0.350	9	36	49	5
20	B	.43	0.136	23	43	28	6
21	C	.69	0.360	6	3	69	21
22	C	.57	0.199	6	3	57	32
23	C	.42	0.118	5	35	42	18
24	D	.57	0.276	22	10	10	57
25	C	.89	0.227	5	2	89	5
26	A	.34	0.077	34	9	10	46
27	A	.81	0.194	81	5	8	6
28	B	.42	0.157	18	42	29	10

The “percent choosing” columns provide 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 performed during test development, and served to identify implausible responses. Continued analysis can inform future revisions of the test. For example, item 13, response D, was chosen only once. Another, more plausible alternative should be considered for the item.

In summary, the test was administered via the web and by pencil and paper to 172 education students enrolled at the University of Central Florida. Test scores were distributed fairly normally, and ranged from 2 to 20, out of a possible score of 22. The mean score for the sample was 11.97, or 54.42%. The Kuder Richardson statistic for internal consistency was .675 with 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.

Validity

All introductory statistics textbooks portray validity, generally defined as determining whether a test measures what it purports to measure, as fundamental to any study. It is also not uncommon to see "validity" represented as something distinct from "construct validity," and often reported both in the scholarly literature and in numerous statistics courses web 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. Clark and Watson (1995) state most succinctly, "Construct validity cannot be inferred from a single set of observations…" (p. 310), but instead offer that a number of procedures should be used to demonstrate construct validity.

Construct validity checks were interwoven throughout the study from inception to analysis, and began with the review of literature. Literature reviews serve to clarify the nature and range of the content of the construct, identify problems with existing measures, and can indicate whether a scale is actually needed (Clark and Watson, 1995). The large number of test items written for Phase I of the study allowed for adequate sampling of breadth of content and representation of a number of items for each content cluster. Subsequent procedures included checks for content validity, factor analysis, and criterion-related validity. Results of each of the procedures are described below.

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 extraneous material has been omitted, but that all relevant material is represented. For this scale, characteristics of the construct of information literacy were represented by the ISTE *NETS*T* standards and ACRL objectives. These criteria describe what content should be included in information literacy instruction, as well as cognitive knowledge students should have 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 construct.

As described in the Methods section, five content experts were asked to evaluate each of the items on the criteria of accuracy, clarity, and institutional objectivity. Averages of reviewer scores for the 22 test items that were included on the ILAS-ED are presented in Table 5. For the rating of item accuracy, reviewers were able to assign fairly consistent ratings across the items. When reviewers were asked to evaluate each item on a scale of 0 (low) to 3 (high) regarding how accurately the item described the objective, all five reviewers scored the items at a level of 2 or 3 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 all the items was 2.47. As the test was devised to be used across multiple settings, institutional objectivity of the item was another important consideration. Using the same 0 to 3 scale as for accuracy and clarity, the experts scored institutional objectivity very highly. All item average scores were 2.2 or more. The mean average for objectivity across all items was 2.85.

Table 5. *Mean Average of Reviewers' Scores, by Item*

Item # (ILAS-ED)	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
Average	2.67	2.47	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 deemed consistently excellent. The accuracy of items as they relate to an identified information literacy learning objective, their clarity, and their institutional objectivity were all corroborated by the content experts.

Factor analysis. In this test, four content clusters were identified from the *NETS*T* standards. The content clusters were hypothesized to be identifying, evaluating, and selecting finding tools, demonstrating knowledge of general search strategies, evaluating and selecting sources, and demonstrating knowledge of legal and ethical practices. Factor analysis was performed to further explore construct validity by investigating the extent to which the content clusters operationally represented unique factors.

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 Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy yielded a value of .689, which exceeded the .50 generally considered adequate for factor analysis. Minimum Eigen values were set at 1.0. Principal component analysis was the extraction method, and as factors were believed to be unrelated, orthogonal rotation was deemed appropriate.

In the initial factor analysis, the researcher limited analysis to four factors with blanking set at .30. Four factors were specified as there were four content clusters. The four-factor solution explained 33.5% of the covariance among the items. Factor 1 accounted for 14.1% of the covariance and consisted of 10 items with loadings ranging from .31 to .53. Although four of the five items in the first content cluster loaded on Factor One, so did items from the other three content clusters. Factor Two likewise contained the items from the second content cluster, with items from other content clusters present, as well. Further review did not reveal any discernible patterns in constructs and the existence of four discrete content clusters was not confirmed.

Because the ILAS-ED is a new instrument, further exploratory analysis was deemed appropriate and additional solutions were conducted. The first analysis, with no specified number of factors, resulted in an eight factor solution. The eight factor solution accounted for 54.3% of the covariance among items, but content cluster analysis did not offer any increased interpretability of the factors. Subsequent procedures were based on seven, six, and five factors. Blanking was set at .30 for all solutions. When factor solutions were analyzed, no apparent logical structure explaining why items clustered on the factors was found. Results of the eight factor solution are presented in Table 6.

Table 6. *Factor Analysis with No Factors Prespecified*

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
IT 27-D	.728							
IT 21-C	.665							
IT 10-A	.467			-.329			.438	
IT 24-D	.327			.322		.439		
IT 23-C	.300				-.539			
IT 8-A		.745						
IT 28-D		.608	-.328					
IT 19-C		.442	.332					
IT 12-A		.402						.448
IT 26-D			.642					
IT 7-C			.543					
IT 11-B			.527					
IT 17-B			.315	.338				
IT 18-B				.738				
IT 16-B				.601				
IT 25-D				.369				-.479
IT 22-C					.663			
IT 14-B					.504			
IT 9-A					-.304	.416	.542	
IT 13-A						.806		
IT 20-C							.707	
IT 15-B								.774

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

Factor analysis of data did not result in anticipated groupings. Claudia Morner (1993) offered a number of possibilities to explain similar results she received when developing a library research skills test for doctoral students in education. Morner first suggested that the five or six items representing the content clusters may be too small and that some content clusters covered a broad range of knowledge. Morner also noted that a larger sample of students may have led to factors loading more consistently on the content clusters. Possible administration of the test to a larger sample, additional items per content cluster, and revisions to the parameters of the content clusters is indicated prior to using the test to diagnose a student's abilities within content clusters.

Criterion-related validity. Criterion-related validity measures are used to determine the performance of the operationalization of the construct, or 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 was concerned with establishing whether the ILAS-ED measured actual information literacy skills. To measure students' abilities to execute the skill in an authentic environment, students were given a test comprised of a subset of items from the ILAS-ED. To distinguish between the two tests, the original 22 item, web- and print-administered 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 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 the mean score, with a range of 36%-86% correct. Each student answered eight items in the library that corresponded to eight items on 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. The eight items selected for the in-library test were diverse in terms of difficulty, ranging from 8.0% to 36% correct, and in item discrimination, which ranged from .08 to .36. The following is an example of one written test item and the corresponding item for the in-library test.

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."

Results from both tests were compared. Table 7 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. This suggests a fairly high correspondence between the tests, which is an indication that the test reflects students' real performance.

Table 7. *Comparison of Scores for 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%)

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 (mainly 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 reflect validity of student performance as it relates to their information seeking abilities.

Reliability

Reliability checks were conducted to measure stability and internal consistency of the scale. 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 the Kuder Richardson 20 formula. Results of the stability procedure and the internal consistency calculation follow.

Stability. Stability was assessed by comparing test scores from the written test with a later administration of the same test. Eleven students took the test a second time. Students were given the same written test form and instructions as before. Approximately two weeks had lapsed between test administrations.

Table 8 summarizes results of the eleven participants. Of the 11 pairs of 22 items, or 232 pairs of items for the test and retest, 172 pairs matched across test administrations. With 11 participants, the mean change was 2.4 items out of 22, 74% of items matched from one test administration to the other. The test-retest results indicated general stability over time.

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 and descriptive statistics for the test are shown in Table 9. Balancing test parsimony with adequate reliability was a key issue for test development. The Kuder Richardson 20 test revealed a reliability of .675. This statistic is in the adequate, but not good, range for reliability, and is not unexpected due to the relatively low number of test items. The average test taker scored 11.97, or 54.42%, on the test. With a standard error of measurement of 1.29, there is a 95% probability that the scores are accurate to 2.58 points, plus or minus. Given the small number of items, the test demonstrates adequate reliability in terms of internal consistency.

Table 9. *Descriptive Statistics for the Test (N=172)*

Mean	11.97
Standard Deviation	3.74
Standard Error of Measurement	1.29
KR 20 Reliability Coefficient	.675

Table 10 displays internal consistency statistics for each of the four content clusters. Kuder Richardson 20 alphas 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 low number of items in each cluster, the simple lack of knowledge of discrete questions rather than the content of the subscale, or that the content clusters are not indicative of a true subscale.

Earlier researchers have hypothesized the existence of any number of subscales or content clusters regarding library or information literacy skills (Morner, 1993; Project SAILS, 2001). The lack of a coherent pattern of correlation among the content clusters validates findings of earlier researchers, who likewise did not uncover evidence of library or information literacy subscales. Extreme caution is advised when relating diagnostic information of the content clusters to test takers.

Table 10. *Internal Consistency of the Four Content Clusters*

Content Cluster	Mean	Variance	Standard Deviation	Number of Items	Alpha Coefficient
A	2.63	1.80	1.34	5	.450
B	3.39	2.18	1.48	6	.433
C	2.91	2.01	1.42	6	.334
D	3.04	1.12	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 existed 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, web searching ability, and intensity of exposure to library instruction with mean score were also analyzed.

A cross tabulation of gender with mean score did not reveal any important differences. Of the 170 respondents who answered the question, males comprised 20.00% of the sample and females 80.00%. The mean score for the 34 males was 11.44 (SD=3.69), and the number of correct answers ranged from 2 to 19. With 136 responses, the mean score for females was 12.05 (SD=3.75), with correct answers ranging from 3 to 20. Table 11 presents a summary of the data.

Table 11. *Mean Scores of Students by Gender*

Gender	Mean Score	Standard Deviation	Number	Valid Percent
Male	11.22	3.69	34	20.00
Female	12.05	3.75	136	80.00
(N=170)				

Ethnicity compared to mean score likewise did not reveal any important differences. Of the 162 students who answered the question, the 140 students who marked their ethnicity as White or European American had a mean score of 12.18 (SD=3.73). The 14 Black or African-American in origin students had a mean score of 10.77 (SD=3.42), the 13 students who indicated Hispanic or Latino in origin had a mean score of 10.92 (SD=3.75), 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. No statistically significant differences among groups were found at the .05 level for gender or ethnicity.

Table 11. *Mean Scores of Students by Ethnicity*

Ethnicity	Mean Score	Standard Deviation	Number	Valid Percent
White	12.18	3.73	140	81.87
Black	10.77	3.42	14	8.19
Hispanic	10.92	3.75	13	7.60
Asian	10.75	5.62	4	2.34

(N=171. The “other” category, containing one response, is not represented here.)

As the test is designed for undergraduate students enrolled in a teacher education program, student classification was limited to the responses of freshman, sophomore, junior, and senior. The relatively fewer numbers of freshmen and sophomores was not surprising, as students are generally accepted into the program after completion of their general education requirements. A summary of statistics is offered in Table 12. Freshmen comprised 8.0%, or 12, of the 150 responses, sophomores 6.7%, or 10, juniors 32%, or 48, and seniors 53.3%, or 80. Twenty-two students did not answer the question. 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.

Table 12. *Mean Scores of Students by Student Classification*

Student Classification	Mean Score	Standard Deviation	Number	Valid Percent
Freshman	10.42	2.75	12	8.00
Sophomore	11.50	3.60	10	6.67
Junior	10.38	3.27	48	32.00
Senior	12.55	3.93	80	53.33
(N=150)				

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. *Mean Scores of Students by Length of Enrollment*

Length of Enrollment	Mean Score	Standard Deviation	Number	Valid Percent
Less than 1 year	10.71	3.49	41	24.26
1 to 2 years	11.51	3.49	55	32.54
3 to 4 years	12.26	3.78	50	29.59
More than 4 years	14.26	3.73	23	13.61
(N=169)				

Statistics were also calculated for level of instruction variables. Test items 3 through 6 posed four different scenarios regarding exposure to library instruction. 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 exposure to instruction levels and are presented in Table 14. Students who have had no library instruction have a lower score than students who have had minimal, moderate, or high exposure to library instruction. Surprisingly, however, students who had received “intensive” instruction had the lowest mean score at 9.07 (SD=4.11), than any of the other categories. This may be attributed to student comfort level with the library and its resources. Perhaps students who felt they needed more assistance continued to seek out instruction.

Table 14. *Mean Scores of Students by Varying Levels of Instruction*

Level of Instruction	Mean Score	Standard Deviation	Number	Valid 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
(N=172)				

Researchers have reported that students tend to overestimate their searching abilities, so two questions were included that asked for students to rate their ability to search library databases and their ability to search the Internet to find information. 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 comparison data are presented in Table 15 and Internet searching comparisons are located in Table 16.

Consistent with the literature, students who were most confident in their abilities to search library databases scored lower on average than students who reported in the "good" or "average" range. Students who considered their library database searching skills as "poor" tended to score the lowest on the test. Mean scores of students in comparison to their self-rated ability to search the Internet showed no consistent patterns.

Table 15. *Mean Scores of Students by Library Database Searching Ability*

Self-Rated Ability	Mean Score	Standard Deviation	Number	Valid 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
(N=172)				

Table 16. *Mean Scores of Students by Internet Searching Ability*

Self-Rated Ability	Mean Score	Standard Deviation	Number	Valid 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
(N=172)				

Overall, analysis of demographic variables did not indicate any important differences among student categories. This uniformity demonstrated that the test measured the information literacy skills levels of participants belonging to several subgroups of education students and can therefore be used with confidence across a number of diverse settings.

Conclusions

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. 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 a scale that has demonstrated validity and reliability, such as the ILAS-ED, will lead to more systematic assessment of instruction and thus more credible reporting in the literature. Understanding how instruction impacts information literacy skills levels is a necessary first step to informing and improving planning, curriculum, and instruction decisions and developing a theory-connected practice of effective instructional techniques.

At the discipline and program level, tests are still in developmental infancy, but their use promises to provide deeper insight into students' understanding of how information unique to their discipline is produced, organized, and disseminated. If other researchers respond to the call from ACRL to develop discipline-specific assessment instruments, the methodology described in the study may serve as a model for information literacy skills assessment initiatives in particular disciplinary areas. Also, as new technologies emerge and foci change, it is hoped the development and validation of the ILAS-ED is only the first of many uses and revisions of the instrument for education.

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 skills levels. How results are analyzed, interpreted, and applied, however, is dependent upon the reason for assessment. While scores can be used to identify an individual student's progress, cohort scores may provide more valuable data for providing a quantitative measure for outcomes based assessment for institutional or accreditation purposes. The instrument can be used for both internal and external benchmarking of education students' information literacy levels.

Different levels of thinking skills are associated with various learning outcomes, and assessment tools should be employed that most authentically measure the skill level. For example, multiple-choice format tests tend to measure lower-order thinking skills, although information literacy emphasizes higher-order thinking processes. The justification for the format 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 ILAS-ED, therefore, is offered as one tool in an information literacy assessment repertoire. Ultimately, no single measure can capture the complexity of learning. To validate an assessment program and successfully measure the range of student achievement, multiple

methods of assessment, administered at critical points throughout the learning process, are necessary. The American Association of Higher Education (2005) writes that learning is multidimensional, integrated, and revealed in performance over time. It is this researcher's opinion that assessment should be, as well.

References

American Association for Higher Education. (2005). *9 Principles of Good Practice for Assessing Student Learning* (Online). Retrieved January 10, 2005 from http://www.aahe.org/assessment/principl.htm.

American Association of School Librarians & Association for Educational Communications and Technology. (1998). *Information literacy standards for student learning.* Chicago, IL: American Library Association.

Association of College and Research Libraries, Task Force on Information Literacy Competency Standards in Higher Education. (2000). *Information Literacy Competency Standards for Higher Education* (Online). Retrieved June 8, 2004, from http://www.ala.org/acrl/ilcomstan.html.

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. (2003). *Characteristics of education doctoral dissertation references: Results of an analysis of dissertation citations from three institutions.* Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

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.

Chadley, O. & Gavryck, J. (1989). Bibliographic instruction trends in research libraries. *Research Strategies, 7*, 106-113.

Daugherty, T. K. & Carter, E. W. (1997). Assessment of outcome-focused library instruction in Psychology. *Journal of Instructional Psychology, 24*(1), 29-33.

Eadie, T. (1992). Beyond immodesty: Questioning the benefits of BI. *Research Strategies, 10*, 105-110.

Edwards, S. (1994). Bibliographic instruction research: An analysis of the journal literature from 1977 to 1991. *Research Strategies, 12*, 68-78.

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.

Grassian, E. S. & Kaplowitz, J. R. (2001). *Information literacy instruction: Theory and practice.* New York: Neal-Schuman Publishers.

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.

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].

International Society for Technology in Education (2000). National educational technology standards for teachers (Online). Retrieved July 19, 2004, from: http://cnets.iste.org/teachers/t_stands.html.

Kennedy, M. M. (1997). The connection between research and practice. *Educational Researcher, 26*(7), 4-12.

Leighton, G. B. & Markman, M. C. (1991). Attitudes of college freshmen toward bibliographic instruction. *College and Research Libraries News, 52,* 36-38.

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.

Middle States Commission on Higher Education (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,* 2070.

National Council for Accreditation of Teacher Education (2002). *Professional Standards for Accreditation of Schools, Colleges, and Departments of Education.* Washington, DC: NCATE.

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.

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 http://sails.lms.kent.edu/index.php.

Rader, Hannelore. (2000). A silver anniversary: 25 years of reviewing the literature related to user instruction. *Reference Services Review, 28*(3), 290-296.

Ren, W. H. (2000). Library instruction and college student self-efficacy in electronic information searching. *Journal of Academic Librarianship, 26,* 323-328.

Schuck, B. R. (1992). Assessing a library instruction program. *Research Strategies, 10,* 152-160.

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.

UCF Office of Institutional Research. (2004). *Preliminary Fall 2004 Headcount by Major and Level*. Retrieved September 3, 2004, from http://www.iroffice.ucf.edu/enrollment/2004-05/1ffal04_allstudents_w.pdf.

Werking, R. H. (1980). Evaluating bibliographic education: A review and critique. *Library Trends, 29,* 153-172.

Zaporozhetz, L. E. (1987). *The dissertation literature review: How faculty advisors prepare their doctoral candidates.* Doctoral dissertation: University of Oregon.