Penny M. Beile, Ph.D., Head, Curriculum Materials Center, University of Central Florida Libraries

The ILAS-ED: A Standards-Based Instrument for Assessing Pre-Service Teachers' Information Literacy Levels

Penny M. Beile

Abstract: Few constituencies exist where it is more important to produce information literate individuals than teacher candidates, yet rarely is it suggested that practitioners newly 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 of teacher candidates. This paper describes the development and validation of the Information Literacy Assessment Scale for Education (ILAS-ED). Test content is based on nationally recognized standards from the International Society for Technology in Education (ISTE) and the Association of College and Research Libraries (ACRL). Procedures designed to enhance the scale's validity were woven throughout its development. This instrument, which consists of 22 test items and 13 demographic and self-percept items, can be used to inform curricular and instructional decisions and to provide evidence of institutional effectiveness for program reviews. The instrument is appended.

Purpose

The purpose of scholarly inquiry is to expand, refine, or refute our conceptual or theoretical understanding of phenomena. 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 suggests 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 recognition of their information need and their

ability to efficiently and effectively meet that 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), the American Association of School Librarians and Association for Educational Communications and Technology (AASL & AECT, 1998), and the International Society for Technology in Education (ISTE, 2000), have 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), thereby implying that the integration of information literacy instruction is the responsibility of all in academia.

Concurrent to these developments, the Association of College and Research Libraries (ACRL, 2000) 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 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 have provided the 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. 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.

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. The purpose of this study was to extend programmatic evaluation of information literacy instruction for pre-service teachers by developing and validating an objective assessment instrument that meets the three-fold challenge of measuring teacher candidates' cognitive knowledge of information literacy, offering usability across unique institutional settings, and providing an instrument specific to the discipline of education.

Perspectives

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 actual 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. Literature surveys of 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). Twenty-seven years ago Richard Werking (1980) concluded that systematic, formal evaluation of library instruction had not occurred to any significant degree; little has changed since that time.

Research indicates several barriers to formal instruction evaluation. Patterson and Howell (1990) offer that many library schools do not offer classes on instructional assessment, thus leaving many librarians feeling they are ill-prepared to properly conduct assessment studies. Eadie (1992) reports that many see formal evaluation as too complex or too time consuming, and may cite the lack of institutional support; he 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 instruction 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; Greer, Weston, & Alm, 1991; and Maughan, 2001) has failed to find instruction positively impacts attitudinal or learning gains.

In every instance the above references studies reported using locally-produced evaluation tools that had not been submitted to validity and reliability analysis. Barclay (1993) acknowledges this limitation, and 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 global assessments.

Attempting to meet this need is the purpose of the Project for the Standardized Assessment of Information Literacy Skills (Project SAILS, 2001), a federally funded initiative which purports to develop 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 and validated approximately 250 general information literacy test items and has recruited a number of institutions to participate 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) call 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 (consisting of this researcher and an Education faculty member) responded to the announcement and was awarded a fellowship to develop an education-specific assessment tool. It is the development, validation, and use of this instrument that is described in this study.

Methods

Scale items were developed based on the ISTE National Educational Technology Standards for Teachers (NETS*T) and ACRL information literacy competency standards. The scale consists of 22 objective, multiple-choice items designed to measure students' cognitive knowledge of information literacy. Expert reviewers rated items for accuracy, clarity, difficulty, and institutional objectivity, while student participants also reviewed each item. A draft of the scale was then administered to 42 participants who closely approximated the target population of teacher candidates. After survey and item revisions larger-scale testing was initiated. Undergraduate teacher education majors at a large, urban university were invited to complete the test. Of the population of over 3,000 teacher education students, 172 completed the instrument.

The test was administered both electronically (n=92) and in print (n=80). Test items were submitted to analysis, and procedures included analyzing distractors for plausibility and calculating item difficulty levels and discrimination indices. A subtest of the ILAS-ED, with ten participants replicating tasks associated with test items in an authentic environment, was administered to check criterion-related validity. Reliability procedures consisted of a test-retest to investigate score consistency and stability, in addition to calculating internal consistency indices. To measure stability, eleven students were administered the same test form twice and results 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

The maximum value on the ILAS-ED is 22 points, and the mean raw score of the sample was 11.97 (SD=3.74). The test has a standard error of measurement rate of .28, which indicates there is a 95% probability that the scores are accurate to .56 points, plus or minus. Table 1 presents descriptive statistics for electronic and print administrations as well as descriptives for the overall sample.

Table 1
Descriptive Statistics for the 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 2 presents summary 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 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 only 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.

Table 2
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

Content Validity

Content validity is generally defined as the degree to which a test reflects all aspects of the dimension or construct being measured. 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, as well as what 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 noted in the Methods section, 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 ILAS-ED are presented in Table 3. For the criterion of 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 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 as 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.

Table 3
Mean Average of Reviewers' Rating, by Item

Item Number	Accuracy	Clarity	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

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 was 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 ILAS-ED. 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*.

Results from both tests were compared. Table 4 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.

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 that changed between the two tests. 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.

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 offer evidence that student performance on information-seeking tasks relates to their test scores.

Table 4
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%)

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.

Stability of the instrument was measured with a test-retest procedure whereby the written test was administered twice. Eleven students were given the same written test form and instructions as they had received earlier. Approximately two weeks had lapsed between test administrations. Table 5 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, Hambleton, & Algina, 1974). Repeated administrations should confirm or negate these preliminary results.

Table 5
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 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 content of the test. 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 they 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.

The Kuder-Richardson 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 .60s and .70s as quite acceptable. 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 thus compromises 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.

Passing Score Calculation

The usefulness of the cut score should be considered in relation to the purpose to which the ILAS-ED will be used. If the test is used to determine mastery of the content and an individual student score is needed to reflect it, 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 calculation 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 cases such as this 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 ILAS-ED 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. 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.

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.

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 ILAS-ED, 76 met that goal.

Assumptions

Two assumptions underlie this study. The first is that test scores are indicative of teacher candidates' ability to locate, evaluate, and ethically use scholarly information. There is question, however, as to whether cognitive knowledge indicates actual behavior. Second, different levels of thinking skills are associated with various learning outcomes, and multiple-choice formats are often thought to measure lower-order thinking skills although information literacy emphasizes higher-order abilities. 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.

Conclusions

Early instruction program evaluators must be applauded for their efforts, yet their results have contributed little to our understanding of instructional efficacy. It is perhaps the lack of rigorously development instruments that has most prohibited systematic investigation of the topic. Rigorously reviewed information literacy assessment instruments are needed to provide information for a variety of purposes at the institutional level. 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 include benchmarking with similar institutions and outcome measures for accreditation purposes.

Skager (1974) reminds us to reflect on the intentions for testing, and then consider the consequences of score interpretation. 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, such as the information literacy assessments offered by Project SAILS and Educational Testing Service, may suffice. However, there is question as to whether education

students are conceptually prepared to take full advantage of the information available to them and testing may be one method used to indicate individual student skills. For these purposes criterion-referenced tests with established passing scores are useful. The ILAS-ED is one example of a test that can be used across purposes, from internal review of instructional efficacy to external benchmarking comparisons to assessing individual mastery. Skager (1974) adds that if objectives are identified and sequenced via curriculum mapping, then the same test can be used throughout the program of study. Given this, the ILAS-ED could be administered repeatedly, with the expectation that scores would increase as students progressed throughout the curriculum.

Ultimately, no single measure can capture the complexity of learning. The justification for a brief, selected-response test such as the ILAS-ED 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 a repertoire of information literacy assessment instruments. 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. When with the American Association for Higher Education, Peggy Maki (2005) wrote that learning is multidimensional, integrated, and revealed in performance over time. Assessment should be, as well.

References

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 (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, 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.

Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research, 56*(1), 137-172.

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.

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.

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.

Ebel, R. L. (1972). Why is a longer test usually a more reliable test? *Educational and Psychological Measurement, 32,* 24-253.

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.

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.

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.

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

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.

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.

Maki, P. I. (2002). Developing an assessment plan to learn about student learning. *Journal of Academic Librarianship, 28*(1/2), 8-13.

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.

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

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.

O'Connor, L. G., C. J. Radcliff, et al. (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.

Popham, W. J. & Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement, 6*, 1-9.

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.

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.

Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion-referenced tests: A decision-theoretic formulation. *Journal of Educational Measurement, 11*, 263-267.

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.

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.

Information Literacy Assessment Scale for Education (ILAS-ED)

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