May 2010

Assessing First-Year Information Literacy Skills

Benjamin Michael LaVerriere  
*Worcester Polytechnic Institute*

Madison J. Dickson  
*Worcester Polytechnic Institute*

Michael James Oliver  
*Worcester Polytechnic Institute*

Follow this and additional works at: [https://digitalcommons.wpi.edu/iqp-all](https://digitalcommons.wpi.edu/iqp-all)

Repository Citation


This Unrestricted is brought to you for free and open access by the Interactive Qualifying Projects at Digital WPI. It has been accepted for inclusion in Interactive Qualifying Projects (All Years) by an authorized administrator of Digital WPI. For more information, please contact digitalwpi@wpi.edu.
ASSESSING FIRST-YEAR INFORMATION LITERACY SKILLS

An Interactive Qualifying Project Report submitted to the Faculty of WORCESTER POLYTECHNIC INSTITUTE in partial fulfillment of the requirements for the Degree of Bachelor of Science by
Madison Dickson
Ben LaVerriere
Michael Oliver

3 May 2010

Approved:

________________________
Professor Arthur C. Heinricher, Advisor

________________________
Christine Drew, Advisor

This report represents work of WPI undergraduate students submitted to the faculty as evidence of a degree requirement. WPI routinely publishes these reports on its web site without editorial or peer review. For more information about the projects program at WPI, see http://www.wpi.edu/Academics/Projects.
# Contents

Executive Summary 1

1 The Question of Assessment 4

2 Literature Review 6
    2.1 History of Information Literacy 6
    2.2 Information Literacy Standards 7
    2.3 Information Literacy Assessment 9
    2.4 Instrument Construction 12
    2.5 Information Literacy Assessment and Education at WPI 14

3 Approaching Assessment 15
    3.1 Student Assessment Vectors 16
        3.1.1 Baseline Skills Test 16
        3.1.2 Confidence Level 17
        3.1.3 Skill Retention 18
    3.2 Program Assessment Vectors 19
        3.2.1 Casual Instructor Assessment 19
        3.2.2 Faculty Collaboration 20
        3.2.3 First-Year Education 20
    3.3 Evaluating the Options 21
    3.4 Developing A New Test 22

4 Instrument Development 23
    4.1 Development Model 23
    4.2 Mapping Standards to Scenarios 24
    4.3 Writing Questions 29
    4.4 Preparing to Pilot 30
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5</td>
<td>First Pilot Administration</td>
<td>31</td>
</tr>
<tr>
<td>4.6</td>
<td>Results of the First Pilot</td>
<td>32</td>
</tr>
<tr>
<td>4.7</td>
<td>First Pilot Summary</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>The Second Pilot</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Purpose of a Second Pilot</td>
<td>37</td>
</tr>
<tr>
<td>5.2</td>
<td>Constructing the Full Test Instrument</td>
<td>38</td>
</tr>
<tr>
<td>5.3</td>
<td>Administration</td>
<td>39</td>
</tr>
<tr>
<td>5.4</td>
<td>Second Pilot Summary</td>
<td>39</td>
</tr>
<tr>
<td>6</td>
<td>Observations on Assessment Development</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>Recommendations for Continued Development</td>
<td>42</td>
</tr>
<tr>
<td>7.1</td>
<td>Validation</td>
<td>42</td>
</tr>
<tr>
<td>7.1.1</td>
<td>Specific Concerns</td>
<td>43</td>
</tr>
<tr>
<td>7.1.2</td>
<td>Exemplary Items</td>
<td>44</td>
</tr>
<tr>
<td>7.1.3</td>
<td>Item Design: A Case Study</td>
<td>47</td>
</tr>
<tr>
<td>7.2</td>
<td>Potential Test Items</td>
<td>49</td>
</tr>
<tr>
<td>7.3</td>
<td>Administering the Test</td>
<td>51</td>
</tr>
<tr>
<td>7.4</td>
<td>Assessment Vectors Revisited</td>
<td>52</td>
</tr>
<tr>
<td>7.5</td>
<td>Reference Guides</td>
<td>53</td>
</tr>
<tr>
<td>8</td>
<td>Conclusion</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>56</td>
</tr>
<tr>
<td>A</td>
<td>Sample Casual Assessment Form</td>
<td>60</td>
</tr>
<tr>
<td>B</td>
<td>Informed Consent Agreement for Participation in a Research Study</td>
<td>61</td>
</tr>
<tr>
<td>C</td>
<td>WPI IRB Application and Approval</td>
<td>64</td>
</tr>
<tr>
<td>D</td>
<td>Pilot 1 Test Instrument</td>
<td>69</td>
</tr>
<tr>
<td>D.1</td>
<td>Pilot 1 Results</td>
<td>75</td>
</tr>
<tr>
<td>E</td>
<td>Pilot 2 Test Instrument</td>
<td>76</td>
</tr>
<tr>
<td>E.1</td>
<td>Pilot 2 Results</td>
<td>80</td>
</tr>
<tr>
<td>F</td>
<td>Web Administration Mockup</td>
<td>81</td>
</tr>
</tbody>
</table>
$C = B \log_2 \left( 1 + \frac{P_s}{N_0 B} \right)$

[23]
Abstract

This project details the construction of an assessment instrument to aid the Gordon Library at Worcester Polytechnic Institute in evaluating baseline information literacy skills among first-year students. The project also presents a selection of other options to fulfill the Library’s need for information literacy assessment in general. Included in this report are conclusions from two cycles of pilot testing and instrument revision, as well as suggestions for further revisions and the continuation of IL assessment at Gordon Library.

Keywords: Information Literacy, Assessment, First-Year, Test Development
Executive Summary

In light of the recent surge of information technology and the requirements of various accreditation groups, Worcester Polytechnic Institute’s Gordon Library should improve its methods of assessing both the information literacy (IL) skills of students and the efficacy of its IL educational programs. Our team of students was assembled and set with the task of developing a plan to achieve at least part of this goal. After reviewing the literature surrounding IL education and assessment, we developed a list of possible methods of either program or student assessment:

- Baseline Skills Test — assess students’ IL skill level directly with a practical knowledge test.
- Confidence Level — assess how well students feel they have learned and retained the material from a particular instructional program both immediately following and some time after having participated in that program.
- Casual/Self Assessment — develop a system for the recording of instructor expectations before and observations after a particular Library program.
- Faculty Collaboration — develop a structured method for collecting observations about student IL skills or learning from faculty who incorporate Library programs into their courses.
- First Year Education — work with extant programs at WPI like Insight (first-year advising and mentoring) or the Great Problems Seminars, as well as with Resident Advisors (RAs) and Community Advisors (CAs), to foster early and effective IL education to first year students.

Through conversation with Gordon Library Staff and project advisors, we chose to develop a baseline skills test targeted at first-year students. Our approach to the design process of this instrument was heavily influenced by the research
of Bonnie Gratch Lindauer [13, 17] and Carolyn Radcliff [22], and by the development process undertaken for the Bay Area Community Colleges Information Competency Assessment Project [25]. Based on these and other resources discussing IL assessment development, we developed our instrument according to the following algorithm:

1. Map standards to performance outcomes — determine a set of learning outcomes that correspond to particular Association of College and Research Libraries (ACRL) standards for IL [1].

2. Write test items — construct test items that each correspond to an outcome-standard mapping pair.

3. Construct and deploy a pilot test — pilot the test instrument to a representative sample of the student population.

4. Review pilot results — consider the results of the pilot and discuss potential changes to the test based on those results.

5. Revise test instrument — using the information gathered from piloting, revise test items and format as appropriate.

6. Pilot again — continue the cycle of pilot-testing and test instrument revision when possible.

Our project’s timeframe allowed us to complete two pilots, with an interesting result from the pilot data itself: students appear to have weaker skills when it comes to the actual acquisition of information, but are generally skilled with evaluating and properly using that information. This result led us to propose another method of bolstering Gordon Library’s IL education programs: quick reference guides detailing the primary elements of the use of data retrieval systems, suggested resources for starting one’s research, and the ever-important reminder to ask a librarian. These have already been implemented, to some degree.

The test as presented here has some known issues, discussed in Chapter 7, and the test would benefit from at least one more revision-pilot cycle before serious deployment. By the same token, our research shows that the test must be pilot-tested again if revisions are made, before being deployed for data collection.

In summary:

- The test instrument developed through this project represents a promising development in improving the state of IL assessment at Gordon Library. The
final 10-question assessment can be found in Appendix E. There are still a number of revisions (discussed in Chapter 7) to be made and tested before deploying the test on a large scale, but we remain confident that the test can produce useful results once these adjustments are made.

- We recommend using either Google Docs Forms or another online survey service to administer the test. These services are free, make the test easy to administer, and make it easier to interpret the results because they collect and digitize the data as a matter of course, unlike a paper based test. The existing WeBWorK system, while it was designed primarily for mathematics assignments, may also be a solution, albeit one that requires more setup time.

- We recommend the following development and deployment schedule:
  - Summer 2010: Revise and pilot-test the assessment based on the discussion in Chapter 7.
  - Fall 2010: Analyze results of the pilot test and report findings.
  - Spring 2011: Revise the test further if necessary, following the guidelines described in Chapter 4, in preparation for deployment on a large scale.
  - Summer/Fall 2011: Administer the revised test to incoming first-year students, possibly repeating the assessment at the end of the first semester.

- If the time and resources become available, we also recommend pursuing any of the other assessment options discussed in Chapter 3. The Confidence Level and Casual Instructor Assessment vectors would likely be the easiest to implement alongside this first-year assessment.
Chapter 1

The Question of Assessment

The past few decades have brought us advanced information retrieval systems and a trend towards digital archiving of everything from research papers to role-playing ephemera [14]. Information about any facet of human knowledge can be retrieved in a few minutes, even the most bizarre and obscure facts. The price paid is accuracy: while it is now a simple task to find reliable information on nearly any topic, it is just as easy to locate inaccurate information. It is now crucial for students — and indeed, all citizens of this information culture — to develop the ability to tell the difference between a useful source and a useless one. In this age of information overload, the need for information education is greater than it has ever been.

As a result, the Worcester Polytechnic Institute (WPI) Gordon Library faces a challenging mission: to promote professional information literacy (IL) skills to the students of WPI. The library’s primary hurdle in teaching IL is that, on the whole, the Library is structured as an information resource and not as an educational department. Without a defined curriculum, and in a school where the lack of course requirements is a selling point, both teaching IL skills and assessing the efficacy of that education is a monumental task. Currently, the only long-standing assessment used by Gordon Library is the test given to students who take the Interdisciplinary Research (ID2050) course in preparation for their Interactive Qualifying Project (IQP), supplemented by occasional surveys distributed on campus.

Gordon Library is not alone: Brasley [5] notes that libraries in general tend to engage in self-assessment methods that are “episodic and haphazard,” a product of the evolving place of the library within the college campus and the evolving state of information access. Even considering data from the ID2050 test as well as anecdotal evidence, there is no longitudinal IL assessment program in place at the
This project seeks to begin the development of a standards-based IL assessment program at Gordon Library, so that the Library can evaluate and focus its IL education efforts to better serve the student population of WPI. Since the possibilities for IL assessment are as numerous as the set of skills that the information-literate student should possess, we begin by developing a set of possible assessment methods that target various aspects of IL education at WPI. Conversations with Gordon Library staff and administrators help us to determine that the most immediately useful of these was a practical skills test designed specifically for first-year students, to be tested before or upon their arrival at WPI. The majority of this report details the development, testing, and revision of that test instrument, along with recommendations for its continued use in the context of IL assessment at Gordon Library.
Chapter 2

Literature Review

2.1 History of Information Literacy

The concept of IL has existed, in some form, since Paul Zurkowski’s 1974 paper, “The Information Service Environment: Relationships and Priorities” [33], in which the author proposed means of achieving “universal” information literacy in the United States by 1984. Possibly encouraged by Zurkowski, the American Library Association (ALA) appointed a Presidential Committee on Information Literacy in 1987, charged to:

“define information literacy within the higher literacies and its importance to student performance, lifelong learning, and active citizenship; . . . to design one or more models for information literacy development appropriate to formal and informal learning environments throughout people’s lifetimes; and . . . to determine implications for the continuing education and development of teachers.” [2]

Despite the Committee’s efforts, when Jeremy Shapiro and Shelley Hughes wrote “Information Literacy as a Liberal Art” in 1996 [24], IL seemed to remain a skill in great demand but in little supply.

Perhaps, Shapiro and Hughes suggested, the deficit of IL skills persisted not because of poor teaching per se, but because of poorly-designed curricula. The authors cast aside “the world of short courses on ‘Getting Started with Windows,’ ‘Surfing the Net’ and ‘Bibliographic Instruction’” in favor of an entirely new curriculum, one designed with an explicit focus on IL in a broad sense. To that end, they identified seven “dimensions of literacy” [24]:
• Tool literacy: “the ability to understand and use the practical and conceptual tools of current information technology...that are relevant to education[,]...work and professional life”

• Resource literacy: “the ability to understand the form, format, location and access methods of information resources”

• Social-structural literacy: “knowing that and how information is socially situated and produced...[and knowing] about how information fits into the life of groups”

• Research literacy: “the ability to understand the...tools relevant to the work of today’s researcher and scholar”

• Publishing literacy: “the ability to format and publish research and ideas electronically”

• Emerging technology literacy: “the ability to ongoingly adapt to, understand, evaluate and make use of the continually emerging innovations in information technology...and to make intelligent decisions about the adoption of new [tools]”

• Critical literacy: “the ability to evaluate critically the intellectual, human and social strengths and weaknesses, potentials and limits, benefits and costs of information technologies”

While the broad curricular reform suggested by Shapiro and Hughes may seem as infeasible as Zurkowski’s universal IL, the multifaceted nature of their definition is not unreasonable. Even the ALA Presidential Committee [2] defined an information-literate person as one who is “able to recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information,” a similarly inclusive skill set.

2.2 Information Literacy Standards

From a practical perspective, no definition of IL can promote actual IL education unless it also provides a way of assessing literacy (or learning). In 2000, the Association of College and Research Libraries (ACRL) published “Information Literacy Competency Standards for Higher Education” [1], which built on the
work of the ALA Presidential Committee and established an explicit set of learning outcomes for IL. These standards have been successfully applied at a national level, and as such they are used frequently in IL research and assessment.

Still, it would be unfair to characterize the ACRL standards as the only such specification. Many other IL standards documents exist, such as those created by regional accreditation bodies, and each presents a slightly different categorization of IL skills. In “Defining and Measuring the Library’s Impact on Campuswide Outcomes,” Bonnie Gratch Lindauer [17] describes the development of a set of universal standards for student learning (including IL) based on the ACRL standards as well as many of these additional standards documents.

Based on her ‘master list’ of standards, Lindauer then developed a set of student-learning outcomes. The first of these, shown below, provides a concise statement of IL as an educational goal.

“All graduates are information literate, prepared to be lifelong learners able to effectively identify, access, and use a variety of information resources; proficient with appropriate information technologies; and able to evaluate and apply information to meet academic, personal, and job-related needs.”

“Defining” these standards and outcomes was only half of Lindauer’s research. In order to fulfill her goal of “measuring the library’s impact” — that is, to assess it — she defined a set of performance indicators for each outcome. Some of these measures, such as the amount of time students spend in the library, rely on existing data while others, such as knowledge based tests and open student surveys, require more time and effort but produce more detailed results.

Just as most regional IL standards tend to concur with the ACRL standards, so too do most universities’ IL standards rely on the regional standards. Thus, Lindauer notes, most universities already have IL standards that are consistent with the ACRL outcomes. Speaking locally, goal 7.8 of the New England Association of Schools and Colleges (NEASC) Standards for Accreditation [18] emphasizes the importance of IL, particularly with regard to a student’s overall education:

**Standard 7.8**

“The institution demonstrates that students use information resources and technology as an integral part of their education, attaining levels of proficiency appropriate to their degree and subject or professional field of study. The institution ensures that students have available and are appropriately directed to sources of information appropriate
to support and enrich their academic work, and that throughout their program students gain increasingly sophisticated skills in evaluating the quality of information sources.”

This goal corresponds with Lindauer’s first outcome, as quoted further above; with WPI’s own goal of “lifelong renewal of learning,” established by the WPI Faculty and Board of Trustees in 1987; and with WPI’s Undergraduate Learning Outcomes, more recently endorsed in 2004 [31, 30].

Furthermore, when the WPI University Outcomes Assessment Committee (UOAC) outlined a plan for assessing WPI graduates in 2005, it included provisions for IL education that agree with the ALA definition (and therefore the ACRL standards as well). Specifically, Outcome 3, Criterion 3 of the UOAC Assessment Plan [28] specifies that students should be able to use technology to “acquire, process, and communicate information,” and the entirety of Outcome 7 supports IL education, stating that “graduates of WPI will be able to make connections between disciplines and to integrate information from multiple sources.” With these common standards in mind, we can now consider the various ways by which to evaluate students’ performance with respect to specific IL outcomes.

### 2.3 Information Literacy Assessment

Assessment of IL skills or education may take many forms. In *A Practical Guide to Information Literacy for Academic Librarians*, Carolyn Radcliff discusses various types of assessment, characterizing them by resource load, target audience, and the level of faculty collaboration, among other factors [22]. These practical metrics allow the would-be assessor to choose an assessment type that is both feasible and appropriate for their unique situation. A multiple-choice test might be most appropriate for assessing the skill level of a large group of students, for example, while a focus group might be the best way to evaluate the effectiveness of a new library program or resource. Ultimately, Radcliff concludes that the most useful assessment programs will likely use a variety of methods.

“The Role of Assessment in Library Education,” by Knight [15], provides a practical demonstration of such a multifaceted assessment program. Her group’s approach included an online post-instruction knowledge test, a pre- and post-test for an in-person instruction session, and a practical skill-testing worksheet. Knight used the results of these assessments to determine, with some degree of precision, which areas of instruction which were less effective. In this way, assessment with
multiple methods proved to be a strong strategy, since it produced useful results with less overhead than a single, prohibitively long assessment.

Sharon Stewart’s approach to IL assessment also promotes variety in assessment methods. While discussing classroom assessment in “Assessment for Library Instruction: the Cross/Angelo Model” [26], she notes that a variety of short, quick assessment techniques — the “Minute Paper”\(^1\) and the “Muddiest Point”\(^2\) exercise being the best-known examples — can be used in the classroom to provide near-immediate feedback and therefore the most educationally useful sort of assessment. In a library context, the “classroom” varies from actual classroom sessions to online tutorials, but Stewart’s assessment techniques can still provide useful data.

A point frequently emphasized by IL researchers is that IL instruction and assessment by the library alone will never achieve the same level of efficacy as an IL education program integrated into the general curriculum. In “Effective Librarian and Discipline Faculty Collaboration Models for Integrating Information Literacy into the Fabric of an Academic Institution,” Brasley [5] discusses methods of integrating IL instruction into the curriculum. In particular, she promotes the “Introduction Model,” wherein students are given IL instruction in their freshman or sophomore years of college. Brasley also warns against relying solely on “on-demand” instruction, such as one-off consults, course workshops, and other limited-audience methods. Instead, she provides a dozen case studies of effective IL-curricular integration, ranging from the general “Freshman Clusters” program at UCLA to discipline-specific examples such as the political science program at the University of West Georgia. In each of these cases, Brasley concludes that programs of this sort are “one clear mechanism for addressing the information literacy needs of current and potential students” [5].

It is also worth noting the similarity between the programs Brasley describes and the Great Problems Seminars (GPS) program at WPI. The GPS program provides an interdisciplinary learning experience for first-year students, including research instruction and assignments. As a result, when the University of Massachusetts Amherst (UMASS) Donahue Institute performed an assessment of first-year learning in the GPS, their findings [27] indicated the students who participated in the program developed stronger IL skills than those that did not. This result supports the active inclusion of IL education in the curriculum, whether through

\(^1\)The “Minute Paper” asks students to spend a minute self-evaluating their comprehension of the skills taught in the program.

\(^2\)The “Muddiest Point” has students identify the topic they had the most trouble understanding.
Similarly, Cameron [6] observes that assessments with actual significance, such as course or graduation requirements, correlate with higher student performance. While this sort of academic importance is highly desirable, Gordon Library is unlikely to be able to effect the inclusion of any IL assessment into the “WPI Plan.” Thus, any assessment program implemented by the Library should be analyzed in a manner that accounts for a possible motivational bias (or deficit) among subjects.

On the topic of obstacles to effective assessment, Donald Barclay identifies a hardship common to librarians and indeed all educators when attempting to accurately and efficiently evaluate educational programs: the complexity of the assessment process. He acknowledges that most librarians value time spent on instruction and at the desk, and that they may see conducting evaluations as overly complex or time consuming. Barclay remarks that “the lack of evaluation of library instruction is a matter of too few resources being stretched too far” [3]. To alleviate this stress, he offers a simple solution: to “do the best evaluation you can with what you have,” using available resources to create a small but effective assessment program. By arguing that minimal ‘hard’ data is more useful than either no data at all or anecdotal observations, he provides a practical answer to the problem of how to enact assessment that is scalable to any library’s resources. With Barclay’s method, libraries with insufficient resources for large scale, in-depth evaluations or standardized tests can still gain valuable information for improvement through a method personalized to the specific needs of the library or institution.

Elizabeth Carter, Associate Professor and Information Services Librarian of The Citadel, a military college in South Carolina, based her evaluation plan for the Citadel’s library on Barclay’s paper. Following his suggestion, she developed an assessment instrument using a variety of methods: a pre- and post-test with multiple choice and short answer questions, a student survey, and focus group discussions [7]. Building on a study done earlier with the Citadel psychology department and Daniels Library, another library on the Citadel campus, Carter designed a test to be administered to first year students during their orientation and again at the end of their first semester. The test was scored by averaging students’ scores on each item, evaluated double-blind by two staff members, and by noting trends in the multiple choice answers. Carter acknowledged that while it was “not as elaborate as a series of t-tests,” the assessment provided data that were indeed useful to the library, since the simplified program was easy to administer and interpret [7]. This information, combined with the results of the surveys and dis-
cussions with a focus group, allowed the Citadel library to successfully tailor its educational programs to meet students’ needs, to develop another, larger assessment, and even to offer a “Library 101” course to students. With these programs, the Citadel library was able to “reach faculty and students in an organized and controlled environment, one where [library staff] can show results” [7].

2.4 Instrument Construction

Much of IL research is devoted to the construction and use of IL assessment instruments. In general terms, Radcliff, Gedeon and O’Connor’s “Assessing Information Literacy Skills: Developing a Standardized Instrument for Institutional and Longitudinal Measurement” [20] describes a number of considerations to guide the development of an IL assessment tool:

- assesses at the institutional level, not the instructor level
- allows for longitudinal data gathering
- allows for pre-testing and post-testing
- is quick and easy to administer
- is a standardized instrument for use at any institution
- is geared toward national standards for information literacy
- assesses both cognitive and affective dimensions

These criteria are not only important for the particular instrument discussed in the article, but represent the desired goals of most IL assessment instruments.

Jessame Ferguson and her team cite these principles as an influence on the development of their IL test for incoming freshmen at the University of Maryland, Baltimore County (UMBC) [10]. This instrument was then used for the first time to assess the IL skills of students at the university in the fall of 2003. Ferguson and her team initially piloted their test to all first-year students. Upon receiving the results of the initial pilot, the team chose to focus their continued assessment on biology students because they provided the highest survey response by a significant margin, which would thus likely provide more accurate results.

The UMBC assessment plan included a student opinion survey asking the importance of IL as one of its first questions. The results of this survey suggested that, while students mainly used the web as their starting point for research, they
felt more comfortable researching from within the library itself. Ferguson con-
cluded that students were not confident in their own abilities, but knew where to
find assistance within the library [10]. The study also found that students were
unfamiliar with many IL skills, including search techniques, determining quality
of sources, and using correct citations in their research.

The Technological Education Institute (TEI) in Thessaloniki, Greece took on
the task of assessing its entire campus to determine whether the school’s curricu-
lum and library seminars have had an effect on student IL skills [16]. This need
arose from the observation that, while the need for IL skills is growing, there
have been no studies of student IL skill level carried out in Greece. The TEI as-
seSSment team developed a 12-question assessment based on the works of several
researchers, including Radcliff, Carter, and Ferguson [22, 7, 10]. They deployed
the instrument to junior students who had had a term each for thesis writing and
practicum within their major, reaching students in all 18 departments of TEI’s five
schools. The results showed relatively low levels of IL skills, particularly when it
came to research strategies, but the authors also note that “it had been indicated
that most professors [at TEI] do not oblige their students to conduct research” [16].
In addition, the study showed disparate score ranges for students of theoretical-
and research-based sciences as compared to practical and “hands-on” sciences,
with the former group scoring higher overall.

WPI, through the Gordon Library, has implemented some IL assessment pro-
gramS of its own. The Library has administered pre- and post-tests both to classes
that use library research services and the ID2050 class. Student feedback collected
by the library [9] indicates that students taking ID2050, offered for their junior
year, recognize that the IL skills they learn are useful, but also that they could
have used the skills earlier in their college careers. Over ninety percent of ID2050
participants either agreed or strongly agreed that research consultations with the
Library helped improve their ability to research for their project. However, there
is currently no system to determine whether students retain these skills through
graduation and beyond. Such post-graduate assessment is beyond the scope of
this project, although the test instrument and insight that result from our work
could serve as a starting point for developing a program to assess the IL skills of
WPI graduates.
2.5 Information Literacy Assessment and Education at WPI

Throughout current IL research, we note two main trends: first, the belief that any form of assessment is better than none, and second, that an understanding of students’ skills and needs allows a library or school to offer the most useful forms of IL education for their population. The state of IL assessment and education at WPI can benefit from both of these conclusions. The current IL assessment methods at Gordon Library are somewhat sparse, and as a result neither the Library nor the WPI administration can quantify the IL skills of the entire student population. Instead, the Library can only make conclusions about particular sets of students, such as those who participate in IQPs abroad. Thus, there is a need for an assessment program that includes all WPI students, and that is feasible with the Library’s resources. Such a program would ideally also include some assessment of the educational efficacy of Gordon Library’s IL instructional programs, since this knowledge would allow the Library to better measure their success in meeting university and regional standards for IL education.

The purpose of this project is to create a solution to resolve this problem. Given the wide variety of possible assessment methods, and the knowledge that any of them could be useful in a custom assessment program for Gordon Library, we began to develop a variety of options for assessment. These options can generally be categorized as assessing either students (their skill level, learning over time, etc.) or education (program efficacy, perceived value of instruction by students, etc.), and are discussed in the next chapter.
Chapter 3

Approaching Assessment

An understanding of the core IL literature, as detailed in Chapter 2, could only help us understand IL education and assessment in general. We sought to also understand the IL education programs offered by Gordon Library and how, if at all, assessment was incorporated into the Library’s efforts. Through several conversations with Library staff, we found that Gordon Library uses several small scale assessments that only reach a small subset of the student population, and also casually measures the efficacy of its programs with assessments in the Cross/Angelo format [26].

One weakness of Gordon Library’s IL programs and assessments may be that they lack a clear connection to specific standards. To address this deficiency, we analyzed the relevant standards documents (particularly the ACRL standards) and compared these with WPI’s own student learning outcomes goals, seeking the common points to serve as a foundation. Interviews with Gordon Library staff also helped identify the specific needs of the Library and of WPI students.

We then developed a set of possible assessment methods that could address a variety of IL education and learning needs at Gordon Library. Each of these “assessment vectors” was designed to provide a different approach to the problem of IL assessment, tailored to various target audiences or educational goals. Because the possibilities for IL assessment are so varied, we developed the assessment vectors with the intent of creating a set of independent solutions, rather than a set of similar solutions for one area of assessment. Here we discuss each vector in detail, considering why the area of IL assessment it addresses is important as well as the particular form it might take at WPI.

It is important to note that these vectors fell into two main categories. Student-assessment vectors rely on direct student interaction and are an assessment of
their skills, while program-assessment vectors focus on evaluating the efficacy of individual Library programs or suggest areas for new programs. There is no single method that can accurately cover the multiple facets of IL, so each of these offer a unique approach to part of the assessment development process.

3.1 Student Assessment Vectors

3.1.1 Baseline Skills Test

By far the most common form of IL assessment is the knowledge (or skills) test. As the name implies, baseline skills assessments are designed to measure students’ proficiency at various IL-related tasks so that a comparison can be made at a later date, usually after a particular instruction program. While there have been some sporadic studies of student IL learning completed at Gordon Library, there has yet to be an assessment of the IL skill level of the entire student body, let alone a baseline assessment made at the beginning of the first year. Armed with such knowledge, the Library could tailor their IL educational programs to address the particular needs demonstrated by the students.

A number of IL tests already exist and are available for use, but the primary issues with these prepackaged tests are expense and lack of flexibility. Of note are the iCritical assessment, Standardized Assessment of Information Literacy Skills (SAILS), and Bay Area Community Colleges Information Competency Assessment Project (ICAP).

iCritical is a test designed by ETS, makers of the familiar SAT and Advanced Placement tests. It is a complex assessment product marketed to universities and made to be easy to use. While it produces targeted individual performance reports for each participant, as well as group statistics, iCritical is rather expensive.

The Standardized Assessment of Information Literacy Skills (SAILS) is a simpler multiple-choice test, and costs three dollars per student with a cap of $2000 for up to five thousand students. Furthermore, the results are processed by the providers of the test and delivered with some basic analysis to the institution. The Gordon Library budget might be able to accommodate this expense, but the scheduling and form of the test is prohibitive considering WPI’s tight academic schedule. SAILS is administered only twice a year at specific times, making it even less feasible for Gordon Library to encourage high participation rates. The test is also web-based, meaning that the Library would somehow need to funnel the entire student population through computer labs within the specified adminis-
tration period. For these reasons, it is not an effective solution.

The Bay Area Community Colleges Information Competency Assessment Project (ICAP) is the product of a team of educators determined to make their own test free for anyone to use. As such, this test does not present the budgetary or logistical challenges of SAILS and iCritical. However, the final version of the instrument was not field-tested, and institutions seeking to use it must first do some work of their own to make sure it is appropriate for their situation. Thus, it is comparable to writing a new test from the ground up, though it may be advantageous to use ICAP as a starting point for a new assessment.

Because the available, ready-made IL tests are inappropriate, we proposed that a test be designed specifically for the temporal and budgetary needs of the WPI campus. Given the opportunity, it should be built from the recommendations and examples of Lindauer [17], Radcliff et al. [22], Barclay [3], and the other authors mentioned in Chapter 2.

### 3.1.2 Confidence Level

While a skills test provides the most rigorous measure of students’ IL skill levels, the resources, time, and coordination necessary to administer such an assessment on a recurring basis make it somewhat less desirable. As such, the idea of “microassessments” designed to measure students’ confidence and self-reported use of various IL skills is attractive.

Such an assessment would be administered in two stages: students would first be asked to self-assess immediately after using a library resource or attending an instructional program, and then to complete a similar assessment three to four weeks later. In both instances, the student would be given a short questionnaire, asking a series of questions about the student’s perceived value of the resource or program.

For the questionnaire immediately following a program, students might be asked:

- Do you feel that you have learned new skills or developed existing skills through this program, and if so, what skills?
- Do you feel the skills you have learned will be useful in your coursework?
- Are there topics or skills you feel were lacking in the program?

For the follow-up questionnaire, students would be asked to reflect on their use of the skills they previously learned:
• Can you still perform the tasks or activities covered by the program?
• Do you feel that the program was worthwhile and useful to you?
• Have you had opportunities to use these skills in actual coursework?
• Have you noticed any gaps in your knowledge since completing the program?

This self-reported qualitative measurement cannot give the Library the precise statistics of a skills assessment. Still, its simplicity and ability to be applied to assess a variety of resources allows the Library to obtain a multifaceted summary of actual student experience, categorized by whatever demographic data is included on the questionnaires.

3.1.3 Skill Retention

The assessment methods discussed thus far are limited to ‘one-off’ assessment. However, as with any survey or assessment, a longitudinal measurement is always preferable to a single data point. A simple way to incorporate longitudinal measurement into an IL assessment program — that is, without committing to repeating a full assessment on some schedule — would be a skills test or survey of senior students, either during or after they complete their Major Qualifying Project (MQP), in addition to the existing ID2050 test. Such a survey could ask students if they felt they were prepared with the necessary research skills to complete their MQPs, and in which areas they felt they struggled. If this type of survey were given to students during their project, it could also prompt students to consider utilizing library resources to improve their research. The results of this assessment would help determine an approximate level of student IL skill (and therefore preparedness for their project), as well as areas in which the Library could offer more or better-publicized educational programs.

A more comprehensive measurement could be achieved by assessing students at any or all of the educational ‘milestones’ in the WPI curriculum, for example, as a first-year student, during the Humanities and Arts project, during the IQP, and during the MQP. Skills tests, confidence level surveys, or any combination of assessment instruments could be administered at these points providing a clear view of students’ progression in IL education.
3.2 Program Assessment Vectors

3.2.1 Casual Instructor Assessment

Instructors regularly consider how well a particular program is working, noting how students respond to instruction and how well they perform on exercises or assessments. However, these important impressions are rarely recorded, leading many institutions to implement what Radcliff calls “casual” assessment [22]. Casual assessment by program facilitators allows library staff to engage in meaningful, documented assessment with very little additional effort. Since this assessment transpires over the course of individual instruction sessions, the instructor need only dedicate a short period of time after the program to reflect on the experience. Gordon Library already performs some assessment of this type, particularly of the forms described by the Cross/Angelo model [26]. However, the format described here would provide a longitudinal perspective on student learning that is currently not available from the Library’s casual assessment program.

A casual assessment form records basic demographic data, such as program date and attendance, as well as more subjective metrics, including the facilitators’ observations about possible adjustments to the program. Such a record of instructors’ observations allows a library to evaluate the efficacy of each educational program, as well as to identify particular areas in which students consistently experience difficulty. A sample form that could be used for casual assessment is shown in Appendix A.

Instructors engaged in casual assessment are asked to identify specific educational goals for the program they teach. By explicitly defining a set of expectations for the students, instructors are able to effectively filter their observations during the program. While teaching, the instructor can then ask the students questions based on the particular goals and topics previously specified. Once the program has concluded, the instructor records key observations from the experience. These may be positive, such as “students completed the ‘truncation search’ activity quickly and with enthusiasm,” or negative, such as “despite a half-hour of instruction, only a few students could identify the proper format of a bibliographic entry.” The power of this assessment vector lies in the aggregation of these data: a monthly or bi-monthly review session would allow Library staff to assess the programs offered during that period of time, and to modify these programs’ curricula if deemed necessary.
3.2.2 Faculty Collaboration

A conclusion found frequently in IL-assessment literature is that the most effective method of promoting IL skills is to integrate IL into the general university curriculum [13]. There are many classes at WPI that require student research, and some of these use a research services overview or tutorial provided by Gordon Library. Depending on the topic and format of instruction, faculty members may be present along with their students. In these cases, the faculty members are asked to complete the “Faculty Attending Evaluation Form” to “help [the Library] determine the value of the library instruction session...and make improvements for future classes” [32]. However, the Library currently lacks any way to assess student IL learning in a classroom context. The professors who teach IL-intensive classes observe the effects of the Library’s programs first-hand, making their observations of students’ performance an invaluable resource. One possible method for acquiring these sort of anecdotes, suggestions, and performance statistics would be to include IL-related questions in the WPI Student Course Reports for the classes in question. The current Course Report allows faculty to include customized questions in addition to the standard items provided by WPI; by collaborating with faculty members, the Library could develop a set of questions to be used in this part of the Course Report. Another method could be to develop a form of recording student progress, whether anecdotally through observation and grades or through a skills test designed in collaboration between the library and the course instructor.

3.2.3 First-Year Education

Helping students become familiar with their library and the available research tools is key to promoting IL skills. Working with first-year students is often cited as an effective way to start teaching IL skills, as in the example of the Citadel library [7]. Actual feedback from WPI students in their junior year [9] supports this belief, with students expressing a desire to have learned research skills before they began their IQP. Rather than waiting for students to come to the Library when the need arises, promoting university-level research skills in the first year would prepare students for smaller research papers as well as major projects later in their education, fostering good research habits.

In turn, to provide the most useful services to first-year students, it is important to understand their research skills and abilities. A basic skills test or survey

---

1See the Library Instruction Request Form at https://www.wpi.edu/Academics/Library/Faculty/instruction-request.html.
administered before or during the students’ first few weeks at WPI would be an effective way to assess the students’ initial skill level. The test would be a modified and simplified basic skills test, similar to the ID2050 survey but tailored to assess an appropriate IL skill set for the first year. These data could be used, in concert with other information collected further along in the students’ time at WPI, to determine whether students who take certain classes or utilize library resources learn more than students who do not, and to assess the relative efficacy of those programs or classes.

Creating an instruction module that Resident Advisors (RAs) and Community Advisors (CAs) can use with their residents could also create an opportunity for first-year IL skill development, providing near-immediate indication of students’ skill levels. For example, a short fifteen to twenty minute instruction period with a library staff member could be followed by an interactive IL quiz activity, requiring the students to use their new knowledge to answer questions in a friendly competitive setting. The feedback this activity would provide depends somewhat on how students are asked to participate: small teams or pairs would probably demonstrate greater overall skill level than students answering independently. The data collected from the scores of each team or student would represent the approximate skill level of that group of students. To this end, a number of schools have hosted the Digital Literacy Challenge [11], an interactive IL skills activity created by a group of Purdue students. Incorporating technology such as the Classroom Performance System Clickers from the WPI Academic Technology Center could remove the need for paper tests and make digitizing their answers more fun than an online survey would be. With high participation rates, a program like this could prove a useful basis for making decisions about IL education.

3.3 Evaluating the Options

We have described the seven potential assessment methods that we developed as the first phase of our project. At this point, because this project is designed to serve the Library, it was appropriate to have the Gordon Library administrators and staff help make the decision of which direction to pursue. After we discussed these options with the Library staff, they recommended that a combination of the baseline skills assessment and first-year student interaction vectors would be the most beneficial direction for the project. An assessment of this type fills in the most urgent missing data in the decision making process of the Library: “what do students know?” It yields information about the skills of students who have
yet to participate in any of Gordon Library’s IL education programs, allowing the Library to target existing gaps in knowledge rather than provide an overview of all facets of IL information. Understanding their skill level helps the library to develop new education programs for first-year students.

### 3.4 Developing A New Test

The Library’s desire for an IL knowledge test presented another opportunity to consider how extant IL assessment instruments could be used at WPI. As discussed in Section 3.1.1, there are a number of tests in use that have been designed specifically to assess IL skill level in college students. However, as was mentioned in the same section, extant tests tend to present obstacles of cost or in flexibility. By developing a new assessment instrument specifically for WPI, we will bypass both of these constraints. In addition, because our assessment is designed explicitly to test IL skills developed in or before the first year, the results can be more accurately associated with first-year students than would be possible with an assessment designed to assess college students in general. A generalized test of that sort could include items testing IL skills that WPI does not expect students to develop in the first year, thus resulting in lower performance overall among first-year students. In contrast, our test should not include any of these “advanced” IL skills, meaning that first-year students can be expected to perform reasonably well on any item in the test, although the placement of the test administration (at the beginning of the first year) may result in some students being unfamiliar with particular skills. It is these skills which Gordon Library’s educational programs should promote to first-year students.
Chapter 4

Instrument Development

4.1 Development Model

We used the ICAP test’s development process as a model to guide our own instrument design, since the ICAP developers have published detailed documentation about the process [13]. The ICAP team first wrote a collection of questions and devised a mapping of student learning standards to question scenarios, then proceeded with a cycle of pilot testing and revision, culminating in the test’s final revision. We chose to reorganize the ICAP methods to some degree, however, based on the retrospective analysis and commentary by the ICAP designers. For example, we found the standards mapping concept a useful tool for ensuring test coverage. The ICAP team developed their mapping only after they had begun writing questions, but it seemed appropriate for us to construct an outcomes mapping before constructing the items for our assessment. By specifying the outcomes to be assessed from the start, we hoped to reduce the number of redundant questions we might construct, and to make sure that items were written to address all of the standards which we hoped to assess.

Another way in which we modified the ICAP development process was related to the piloting and revision cycle. The final published version of ICAP was never pilot-tested, for reasons discussed in Gratch-Lindauer & Brown [13]. As such, institutions considering ICAP are warned that “[t]he exam was not sufficiently field-tested for the Project Team to confirm its absolute validity and reliability” [25]. While the previous pilot testing should guarantee that this final, untested version of the assessment is still relatively accurate, we felt it was wisest to structure our development process to ensure that our final instrument had been pilot-tested.
For this reason, the final instrument (in Appendix E) does not reflect a number of modifications that would probably improve the test. (These changes are discussed in Chapter 7.)

The overall pattern of instrument development, including our modifications, is illustrated in Figure 4.1. First, standards and criteria must be defined for the assessment. This stage includes the work done in researching IL and university standards, and in devising our outcomes mapping. Second, we delve into the actual design and development of the items for the test, paying attention to the mapping constructed in the previous stage. Based on the items, we create a test specification document (shown in Appendix G), noting the rationale for each item, as well as for test-level decisions like length, administration method, and item format. Finally, we enter a testing revision cycle where the questions and format of the prototype assessment are pilot-tested, the results are reviewed, and the instrument is revised.

4.2 Mapping Standards to Scenarios

As discussed previously, the first stage of the test design process was to specify clearly the test’s purpose based on the official standards we had chosen: those of the ACRL. In the vein of Barclay’s “do the best evaluation you can with what you have” [3], we felt it would be most appropriate to select a smaller group of standards — and be confident that we could assess them well — than to attempt the monumental task of designing a comprehensive IL assessment program. Based on our discussions with Gordon Library staff and administrators, we chose to assess standards that represented the IL skills which we expect students to learn before or during their first year at WPI. In addition, we attempted to select standards that could be assessed through a written test in multiple-choice format. The format of our assessment, a single test administered to students individually, naturally excluded certain ACRL standards. In particular, standards that focused on group work and practical project-based skills were excluded because of the complexity and work involved with facilitating such an assessment on that scale. Also, because Gordon Library has limited staff, time, and resources to devote to assessment, it would have been unwise to choose questions of a format that could not be tabulated easily, such as essay or ‘short answer’ items. While it has been noted [3] that multiple choice questions might not emulate the open ended nature of researching in a library, they are a commonly used assessment method and the data returned from these questions are still useful. After considering the Institution’s
Figure 4.1: Test development process, based on Gratch-Lindauer & Brown [13].
goals, the test’s target audience, and the other factors discussed above, we selected the following list of standards as those that are reasonably assessable (from American Library Association [1]):

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

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

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

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

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

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

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

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

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

3.1.a Reads the text and selects main ideas.

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

5.1.a Identifies and discusses issues related to privacy and security in both the print and electronic environments.
5.1.b Identifies and discusses issues related to free vs. fee-based access to information.

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

Next, we developed a set of IL task related scenarios that we felt students should be able to accomplish during their first year at WPI. These were cross-referenced with the list of standards to produce a standards-to-outcomes mapping, shown in Table 4.1 on page 28. Each column of the matrix corresponds to an ACRL standard, and each row represents one of the ‘real-world’ scenarios we developed. A number in a cell indicates that an item designed to test the scenario corresponding to that row is a way of assessing the particular ACRL standard for that column. Not all scenario-standard pairs are sensible, let alone assessable; there are also a number of pairs that could be assessed but do not correspond to any item in our assessment. These potential new items are discussed in Chapter 7.
<table>
<thead>
<tr>
<th>Task</th>
<th>1.2.e</th>
<th>2.1.c</th>
<th>2.2.*</th>
<th>2.3.b</th>
<th>2.5.c</th>
<th>3.1.a</th>
<th>3.2.a</th>
<th>5.1.d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtain resources by a variety of methods</td>
<td>6</td>
<td>7</td>
<td></td>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construct search strategies</td>
<td>c</td>
<td>a</td>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluate source type, relevance, and validity</td>
<td>6, 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1, 6, 7, 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gather bibliographic data and cite resources correctly</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4, 10</td>
<td></td>
</tr>
<tr>
<td>Read and evaluate sources critically</td>
<td>e</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understand and avoid plagiarism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

Table 4.1: Students’ ability to perform the task listed in each row should correspond to mastery of the ACRL standard listed in the column header. Thus, an item number in a cell indicates an assessment item designed to test the corresponding task outcome. (Item numbers refer to those ultimately used in the second pilot, as shown in Appendix E.) A cell containing a letter refers to a potential new question, as discussed in Section 7.2.
A few minor adjustments were made to our original list of “assessable standards” during the construction of the standards mapping. First, the scope of ACRL standards in the 2.2.* range were deemed similar enough to be included as a single column. Second, although we originally listed the 5.1.a and 5.1.b standards as assessable, further consideration of the standard and discussion with our project sponsor led us to conclude that these standards fell outside the scope of skills we wished to evaluate and could be discarded from our assessment. The legal and moral issues of the use of information are too complex a topic for a short assessment of basic IL skills.

The standards-to-outcomes mapping has threefold importance for the instrument-construction process. First, it guarantees that each question has a purpose, as each is modeled after a specific task linked to one or more ACRL standards. This method of item construction is more efficient than the process used to construct ICAP, since each item produced necessarily corresponds to a particular standard. As a result, there should never be a need to discard irrelevant questions or write “replacement items,” a problem that can arise if items are constructed independent of an outcomes mapping. Second, the mapping plainly shows if there is a standard or scenario that is unnecessary or untestable, allowing, if need be, the refactoring of the initial goal. Finally, the mapping provides a quick benchmark for deciding between alternative versions of questions while reducing the number of unnecessary duplicates, or questions that assess the same standard or scenario: if two questions address the same set of scenario-outcome pairs, one can possibly be eliminated. Not all “duplicates” in the mapping should be removed, however. The 2.2.* range, for example, can include a variety of searching techniques, leading to a set of questions that occupy a similar range in the mapping but which actually test an important set of distinct skills.

4.3 Writing Questions

With a mapping of real-world scenarios to the ACRL standards, we could start writing questions. As items were constructed, they were either designed from or checked against the outcomes mapping to ensure that all scenario-standard combinations specified had been accounted for. Guiding question design in this manner proved to be effective at both quantitatively covering the skills we wanted to test and reducing the number of questions generated and retained for the test itself.

In a few cases, we adapted or used verbatim a test item from the ID2050 test already in use at Gordon Library. When revisions were made, it was generally
to improve the format of the item based on our research into test design “best practices.” As such, a secondary result of this project is to provide these revised items as examples of improvements that could be made to the ID2050 assessment itself.

Finally, the initial collection of test items was reviewed for clarity of phrasing and expected task. Poor vocabulary choice, incorrect assumptions, and several other mistakes were noted and edited. The end result of this stage was a short list of eleven questions covering the scenario-to-standards mapping neatly and succinctly.

4.4 Preparing to Pilot

With a set of test items in hand, the next step was to test on a representative population. It was important that the first round of piloting focused on the quality of the questions themselves and not on the performance of the subjects. We first wanted to test whether items were understandable, were of average skill level, and actually tested the areas each was intended to measure. To gather this data we decided to discuss the questions for a few minutes with each participant after completing their test.

The goal is to assess the test questions, not the students, so the actual scores obtained through the first pilot are to help in grading the questions against the criteria described above. For piloting, a short survey was added to the test instrument, allowing us to analyze the data with in the context of some basic demographic data about the subjects. Specifically, the survey requested subjects’ major, class year, and previous classroom or library IL-education experience.¹

When recruiting subjects for the pilot test, we restricted participation to first- and second-year students only, so that subjects would possess a level of IL knowledge similar to that of our target audience. We hoped to have a subject group of around thirty students, although even a small number of students would allow us to identify the majority of problems with the test instrument [19]. For a group of this size, it was easiest to deploy the test on paper; this choice also allowed us to take notes on our discussions with subjects on the test instrument itself. To encourage participation, particularly among the survey-besieged population of WPI students,

¹In retrospect, this survey should also have asked if the subject had participated in the Great Problems Seminars, a first-year program with an emphasis on research skills. When the UMASS Donahue Institute studied the program, they found that student awareness of research skills and tools was greater among GPS students than among students who were not in a GPS class [27].
we appealed to that oldest of motivations: free food. The offer of free brownies for participation in the pilot proved reasonably successful, although other factors (discussed in Section 4.6) may have limited the tactic’s effectiveness somewhat.

Since a short discussion needed to follow each pilot, and the importance of the subject’s impressions on each question needed to be acquired, testing the whole set of questions was deemed inappropriate. Instead, four subsets of three to four questions were compiled into four different pilots\(^2\), utilizing alternative versions of questions to differentiate the better written ones. Participants could easily finish four multiple choice in the target administration time (15–20 minutes) and be able to remember their reasoning and specific feelings for each item.

The items on each of the four tests were chosen in order to cover a variety of subjects and avoid similar questions. Because one question which was particularly long, the test that included it had only three questions for reasons of fairness and time. The four pilots were labeled with animal names instead of numbers or letters to make the test more fun and enjoyable, while making each test unique and identifiable for later discussions.

4.5 First Pilot Administration

Before we could pilot-test the instrument, we were required to obtain approval from the WPI Institutional Review Board (IRB) to perform research involving human subjects. Because the pilot test procedure was part of the development of an educational assessment, and because we would not retain any personally-identifying information about the subjects, our research was deemed “exempt from further IRB review and supervision” according to the U. S. Code of Federal Regulations (45 CFR 46.101(b)(2)) and WPI policy. (See Appendix C for the full IRB application and letter of approval.)

When a subject arrived to participate in the pilot, a facilitator explained the test’s goal and the experimental procedure, which was also detailed in the informed consent form the subjects were required to read and sign affirming their understanding. (See Appendix B.) Subjects were then given one of the four test forms and an answer collection sheet.

After the subject had completed the test, a facilitator would ask the student to compare a certain question against its alternative(s). The facilitator then discussed with the subject his or her reasoning behind their answer for each item, looking

\(^2\)These four tests can be found in Appendix D
at how the subject arrived at their choice and how the subject ruled out incorrect answers. It must be mentioned that nothing rigorous occurred in the post-pilot discussions, and the notes mainly consisted of a single sentence such as “I am a Psych major; everything we do is with journals, so I think that that is the most reasonable source.” Despite its simplicity, this practice seemed to be an effective way of identifying confusion on the part of the subject.

4.6 Results of the First Pilot

In the end, the first pilot was only administered to 18 students. The combined causal factors were timing of the pilot (the end of a term, close to finals) and a constant flow of surveys on campus competing for students’ time and attention. (In this respect, offering brownies as an incentive was a good decision.) Our remote location in Kaven Hall, the academic building furthest from freshmen residence halls, and minimal advertisement may have also hindered participation. However, the small number of subjects does not discredit the value of the results.

The most useful data gleaned from the first pilot were the responses from subjects about their answer choices and thought processes while taking the test. From the discussions with pilot facilitators about how the participant eliminated incorrect answers and their interpretation of the questions, a clear picture of the major issues with many questions developed.

Some items were poorly phrased and subjects interpreted the questions or answers inconsistently. For example, question 2 from the “Ostrich” test read as follows:

**Ostrich, Question 2:** For a 20th century history class, you need to find an interview of someone’s firsthand account of a riot that happened 20 years ago in Chicago. Which of the following would provide the most appropriate results?

A. newspaper database  
B. a journal  
C. Google search  
D. magazine database

Answer B, “a journal,” was interpreted by some subjects as a personal journal (that is, a diary), while the intended reading was as a scientific or peer-reviewed journal.
Another example of unnoticed ambiguity was question one from the “Penguin” test:

**Penguin, Question 1:** You are researching the spread of influenza in the Middle Ages. Which set of keywords will give you the most relevant results?

A. influenza, Middle Ages
B. influenza, flu, epidemic, Middle Ages, medieval
C. influenza, disease, illness, information, research, study, medieval, Middle Ages, statistics

The correct answer was B, but only two participants chose it. We originally wrote this question without considering that subjects might have trouble because they were not aware of how the search system in question worked. Search engines can vary significantly: research databases often use explicit Boolean operators, whereas Google and the like use an implicit OR Boolean operator that employs ‘fuzzy logic’ when compiling results. The question as written was aimed at a generic, ‘friendly’ search engine that would find the best match to the keywords given, in the manner of Google or other general-purpose search engines. Without this information, the question is open to interpretation as to the specifics of the query, which can lead to confusion. Facilitator discussions revealed that this was indeed the case. One subject asked what the implicit Boolean operator was for the query, suggesting that not every subject has the same concept of a ‘default’ search engine behavior.

The second item on the “Penguin” test led to some confusion.

**Penguin, Question 2:** If you were writing a paper, which of the following is the appropriate way to incorporate the quoted passage into your writing?

“Semiotics can be conceived of either as a unified theoretical approach to the great variety of systems of signification and communication, and in this sense it constitutes a metalinguistic discourse dealing with any of its objects by means of homogeneous categories, or it can be conceived as a description of those various systems insisting on their mutual differences, their specific structural properties, their idiosyncrasies — from verbal language to gestures, from visual images to body positions, from musical sounds to fashions. It shows a
A wide range of ‘languages’ ruled by different conventions and laws. It can investigate those various domains either at the elementary level of their consecutive units (such as words, color spots, physical formants of sounds, geometrical or topological shapes) or at the more complex level of texts and discourses — that is, narrative structures, figures of speech and so on."


A. Semiotics can be conceived of either as a unified theoretical approach to the great variety of systems of communication, or as a description of those various systems insisting on their mutual differences. It can investigate those various domains either at the elementary level of their consecutive units (such as words, color spots, physical formants of sounds, geometrical or topological shapes) or at the more complex level of texts and discourses — that is, narrative structures, figures of speech and so on.

B. Umberto Eco defined semiotics as “a unified theoretical approach to the great variety of systems of signification and communication.” He suggests that there is a wide range of “languages” that contribute to communication, both at the elementary level (words, color spots, etc.) and at the more complex level of texts and discourses.

C. The field of semiotics explores the ways in which forms of communication are given meaning by human beings. Umberto Eco suggests that this investigation can occur “either at the elementary level of their consecutive units (such as words, color spots, physical formants of sounds, geometrical or topological shapes) or at the more complex level of texts and discourses — that is, narrative structures, figures of speech and so on” (Eco, 108).

The question, as shown above, required subjects to read a lengthy selection of complex text, and then to choose a valid citation from three choices. Some participants took the meaning of the question to be “what is the best way to write this quotation into a text?” — that is, what is the best choice stylistically to use the information provided in a paper. In these cases, participants completely ignored
the issue of citing the work properly, which was the designed purpose of the question. In addition, while the question does not require any understanding of the text itself, we found in practice that the difficult text only distracted from the question. As such, a more readable passage of text was used when this question was revised for the second pilot.

A particularly egregious error in item construction concerned item three on the “Iguana” test.

**Iguana, Question 3**: Who can access the Library’s information databases?

A. students and faculty  
B. faculty and alumni  
C. students and alumni  
D. students

Student responses revealed that the correct answer was derived from the choices themselves or guessed using simple deduction, and all test takers answered it correctly. Because of an oversight on our part, this question was not modified for the second pilot, and so further discussion of the item can be found in Section 7.1.1 on page 43.

Overall, the comparison of alternate item forms proved to be of little use. Subjects were generally indifferent as to preference of alternates, and these were resolved purely by our preference for a particular wording. Questions involving critical thinking about resources fared better than search and resource questions, with all participants understanding what the questions were asking and most showing some skill in identifying trustworthy resources. This trend extended to questions about selecting the most appropriate resource from a selection.

### 4.7 First Pilot Summary

In developing the test, we modeled the methodology of the creators of the ICAP test. This involved first developing a selection of items based on best practices found in the literature and a mapping of targeted standards and performance outcomes. Following this was a cycle of piloting and revision, with the project finishing after a pilot to provide a known level of validity for the assessment. The first pilot was devised to evaluate the test items themselves, checking for consistency and answerability. To this end, subjects’ thoughts on the questions were
recorded. A variety of problems with the items were identified from this data, and the alterations made to the items themselves are discussed in the next chapter.
Chapter 5

The Second Pilot

After reviewing the results of our first pilot, we turned our attention to rewriting unsatisfactory questions, then designing and performing a second pilot of the test. This time, we were more focused on simulating an actual test environment, but still noted how individual test items performed.

5.1 Purpose of a Second Pilot

The first pilot was a study in item utility; can the questions be relied upon to measure what they are supposed to and not measure how proficient students are at decoding the question? Students’ actual answers were less important than how students interpreted the questions, and the test revealed a few important areas needing revision.

The purpose of the second pilot was to measure the test performance in a more formal context. This necessitated a pilot in the form of the final assessment with all questions included.¹

A more representative population was essential for this pilot and so we chose the group of late-admits, transfer students, and exchange students that were to arrive at WPI for a new student orientation program in January. This was the closest to our intended population for the final test of first year students.

¹Questions were revised from the data collected from Pilot 1, before being included in the second pilot. Alternate versions of questions were combined by taking the good qualities of each and rephrasing for clarity.
5.2 Constructing the Full Test Instrument

Most of the work for designing the second pilot included analyzing the previous results and resolving outstanding issues revealed through the test takers’ interpretations of the questions. The method and reasoning behind many of the changes between the first and second pilot tests are outlined below, and the full pilots can be found at Appendices D and E, respectively.²

The question regarding validity of sources for information of “Bardism” or “Neo-Paganism”, we chose to use the latter on the grounds of realism, and refined the question further, to more clearly define what kind resource we are asking for; specifically a reliable source for a research project.³

The second question from the Penguin test went through a considerable change. After deciding that the quote was too long and complicated for our purposes, it was replaced with a shorter, more comprehensible selection of text. Our revision also switched the word “incorporate” to “quote” in an attempt to better express the point of citation without calling on it by name.⁴

For the question about influenza in the Middle Ages, neither version was appropriate due to the ambiguity of the type of search engine (as outlined earlier). We added “using a search engine like Google” to help correct this problem⁵.

There were two similar questions about sources, one about a firsthand experience of a riot and one for researching a species of plant, that had almost the same multiple choice answers and tested for the same outcomes. We kept both of them because they required different answers, and to differentiate between the two was more important than the individual questions alone. In the question about plant research, expecting ”good results” was expanded into ”relevant and reliable information” regarding sources for clarity.⁶

For Boolean search strings, we tried multiple variations and formatting of the question, and after testing the two we concluded on, we ended up merging the two together. Keeping the question in-line instead of bulleted made it easier to read so we chose the more concise wording, as it was somewhat confusing for students.⁷

On the question for recognizing proper citation styles, we had to choose between “which of the following is” and “which of the following is not” a valid

²The following are question sets of corresponding questions found in the first and second pilots.
³Ostrich: 4, Iguana: 1; Pilot 2: 1
⁴Penguin: 2; Pilot 2: 4
⁵Penguin: 1 ; Iguana: 4; Pilot 2: 2
⁶Zebra: 1; Pilot 2: 7
⁷Zebra: 3, Iguana: 2; Pilot2: 9
citation style. While one of the ICAP question design principles states that positive phrasing tends to be less confusing, we felt that the question became too difficult if we asked the former, since it requires students to know a particular formatting style, rather than recognize one that is lacking.8

5.3 Administration

We were given the opportunity to administer this second pilot to the group on Tuesday, January 12th at their 9am library instruction session in Anderson Lab B. Of the thirty registered students, seventeen arrived on time for the session, while three students were too late to start the test. Each student was given a copy of the full test instrument, containing a set of demographic questions and the complete item set. The full test is shown in Appendix E. One of the demographic questions on the test revealed to us one of the revisions necessary for the next version of the instrument (discussed further in Section 7.1): we had neglected to include “exchange student” as a category into which students could classify themselves.

5.4 Second Pilot Summary

The purpose of the second pilot was to try the test in a more realistic testing environment and format. This involved compiling all of the questions into a single test and administering it to a group of first-year students at the same time in a classroom. The revisions to the items were made using the results of the first pilot, with items that had alternate versions selected for clarity of phrase, realism, or response from the first pilot. Pilot administrators proctored the test just before the second semester in the Library during an information session for new students.

---

8Ostrich: 3, Zebra 4; Pilot 2: 10
Chapter 6

Observations on Assessment Development

After much research and development, we have a test of IL for first-year students. It is short, quick, and a good indicator of general IL skills that can serve as a useful tool in guiding the Gordon Library’s strategies for educating WPI students.

More than a test, though, we have produced a method for test development that is extensible and reliable. When constructing an assessment, it is first necessary to compile the educational standards it seeks to measure in the testing population. These are usually already known, but in some instances a custom list may have to be built. Given the standards that apply, any that are infeasible due to budget, time, or available resources must be excised from the list. For instance, we did not test subjects on their group work skills as there was no way to do so within the confines of the intended format: a multiple choice test.

After determining the standards to be assessed, a second list of demonstrable scenarios that the subjects are expected to be able to do is created. This list is then cross-referenced with the standards to make reasonable combinations of standards and scenarios. This matrix is then used to guide the creation of actual test items, with each intersection of standard and scenario representing at least one question, ideally several. Items can be of any format that is conducive to demonstrating the scenario they represent. After this step, the development moves into a (continuous) cycle of piloting and revision. It is very important in the early stages of the piloting to ask pilot participants how they interpreted items and if possible how they arrived at the answers they gave. This step alone easily eradicates a large portion of errors in item construction that were not obvious to the designers. Finally, never start using an assessment without having piloted the version to be deployed
first.

Our use of the process itself has even yielded some interesting, though not conclusive, results. The performance of students on the second pilot, culling erroneous questions and combined with the reasoning data of the first pilot, suggests that students know more about information evaluation than they do about information retrieval. Overall, students did less well on questions that asked the subject about finding information\(^1\), either through a search engine or a database, than on questions in which the subject needed to, for example, evaluate a source’s validity.\(^2\) This pattern of performance makes sense. Internet search engines are designed to get relevant results for everyday queries and are much easier to use, though not as specific, as those used for academic research. Conversely, students probably had to contend with the evaluation of information frequently in their high school courses.

Though our methods are sound, our pilot tests did not provide enough data for confidence in any conclusion. Because our results lack certainty, not least of all due to small pilot-testing groups, we suggest that the Library first make the sort of revisions discussed in Section 7.1, followed by another pilot on a somewhat larger scale. Once the test has been piloted, it can be deployed however the Library desires, with the knowledge that the results will be a relevant indicator of first-year student IL skills.

\(^1\)Pilot 2 questions 5, 7, and 9 — Appendix E
\(^2\)Pilot 2 questions 1 and 8 — Appendix E
Chapter 7

Recommendations for Continued Development

The process of test development may never end: an assessment can always be improved or adjusted, whether to increase the accuracy of the results or to adjust the test to assess a new population or topic. In this chapter, we discuss a variety of ways in which development of our test instrument can continue.

7.1 Validation

Test design is a complicated business, and the process of validating a test consists mainly of piloting and revising it repeatedly. The timeframe of our project only allowed us to administer two pilot phases, and so the test as it stands now has some flaws and peculiarities that need to be addressed further. This is not to say that revisions should be made as soon as possible, or that revisions should be made after each use of the test: the test cannot function as a longitudinal assessment tool unless the data is collected in exactly the same manner for each instance. If anything, once the test is deployed, it should be used annually for some time while development and piloting of a revised version occurs independently. After a sufficient amount of data have been collected (an amount to be determined by the Library), the revised instrument can be deployed from that point forward.
7.1.1 Specific Concerns

There are a few items in the current test instrument that warrant discussion. These specific issues are derived from the results of the second pilot and discussion with advisors. In short, questions 2, 3 and 5 need to be revised, either in writing better multiple-choice answers or reconsidering the entire question.

**Question 2**: Who can access the Library’s information databases?

A. students and faculty  
B. students, faculty, and alumni  
C. students and alumni  
D. students only

It should be noted that this question does not in fact, align with any of the standards that we intended to assess, as it is institution-specific. We felt it was important to ask this question, though, because students should be aware that access to good research information is rarely cheap and often restricted.

Each answer for question two includes the option ‘student’, which may hint that the answer is not ‘students only’. Furthermore, though Gordon Library does have an official policy on the use of its databases by alumni that prohibits such use, it is a poorly communicated and sometimes insufficiently enforced one. In order for it to remain and be a valid question, the Library must define and communicate its policies more clearly.

Another issue with this problem is that it is has some aspects of a “teaching” question rather than an assessment question, which is in violation of our design specifications. In other words, this question tends to act as an educational item for students, informing them of a particular piece of knowledge, rather than an assessment item that tests existing knowledge. An incoming first-year student would likely not know the answer, but they may learn what the answer might be, and therefore gain knowledge, through answering this problem. We felt the importance of the subject matter out-weighed the exceptions that had to be made to include it.

**Question 3**: Which of the following sets of keywords would be most appropriate if you were searching for information on the spread of influenza in the Middle Ages using a search engine like Google?

A. influenza, Middle Ages
The uncertainties students experienced with this question, in retrospect, could be suggested by the fact that performing the requested searches with Google yields unclear results. Furthermore, participants reliably choose between the first two options with almost perfect randomness. We did not test this question on Google itself since it was not originally written with any specific engine in mind. Without getting into the details, the simple solution to fixing this question is to change the subject matter to a carefully selected and tested one. A complete discussion of the reasons behind this are given in Section 7.1.3.

**Question 5**: A truncation search for “nanotech” in the library catalog will include results related to which set of terms?

A. nano, nanos, nanosized, nanobit  
B. nanotechnology, nanotechnological, nanotechnologies  
C. nanostructured, nanoparticles, nanophotonics  
D. nano-micro, nanoscale, nanocrystalline

Question five has the simple problem that all the wrong answers are wrong in the exact same manner: they do not contain the word ‘nanotech’. This opens the question to guessing by the test taker. The solution is to add wrong answers that are wrong in different ways, such as ‘nanoscale technology,’ which is lexically related to the word ‘nanotech’ but is “wrong in a different way.”

### 7.1.2 Exemplary Items

Despite the lengthy list of potential revisions to be made on our assessment instrument, we feel there are some items which serve their purpose rather well. In particular, the test items discussed in this section possess no obvious flaws, and — while more intensive pilot testing may reveal some slight adjustments to be made — obey most principles of good test item design. These are questions 1, 4, and 6.

**Question 1**: Research scenario: You need to research a religion called Neo-Paganism. Which of the following sources would be most credible for a detailed research project?
A. “Neo-Paganism”, from the Wikipedia article at www.wikipedia.org/wiki/Neo-Paganism
B. “The Pagan Diaries”, an editorial from the local newspaper
C. “Neo-Paganism: from Humble Beginnings”, an article from the journal Philosophy and Religion Quarterly
D. “The Truth about Paganism”, from the website www.thesecrettruth.net/revealed.html

We think this question tests the subject’s skill at evaluating Internet sources very well. It specifies that the situation is hypothetical, clarifying a point which was confusing for some pilot participants during the development process. The answer choices include a powerful distractor: the Wikipedia page. Even some Library personnel admit Wikipedia can be a good place to begin in order to get an overview of a subject. We countered this possibility by specifying that the scenario concerned a detailed report on the subject matter.

Question 4: Read the following text and answer the question at the end.

“Worldwide during 1995, 2.7 million new adult HIV infections occurred (averaging more than 7,000 new infections each day). Of these, about 1 million (an average of nearly 3,000 new infections per day) occurred in Southeast Asia, and 1.4 million were sub-Saharan Africa. The industrialized world accounted for about 55,000 new HIV infections in 1995 (nearly 150 new infections per day: about 2 percent of the global total).”


If you were writing a paper about the spread of HIV/AIDS in third-world countries, which of the following is the most appropriate way to quote the author’s ideas above into your writing?

A. 2.7 million new adult HIV infections occurred (averaging more than 7,000 new infections each day). Of these, about 1 million (an average on early 3,000 new infections per day) occurred in Southeast Asia, and 1.4 million were sub-Saharan Africa.
B. Chris Beyrer writes that in 1995, “2.7 million new adult HIV infections occurred”, meaning there were more than 7,000 new infections each day. Of these, “about 1 million” occurred in Southeast Asia, and 1.4 million were in sub-Saharan Africa.

C. In 1995, new adult HIV infections averaged “more than 7,000 new infections each day”, with about 3,000 of these occurring in Southeast Asia. Further, the unindustrialized world accounted for about 98 percent of new HIV infections in that year (Beyrer, 3).

This item, a streamlined version of a particularly troublesome question in Pilot 1, is now simple, straightforward question modeling a real-life research scenario. Though the content of the text is real, it is inconsequential to the core question: how should a writer cite this? The giveaway of using the verb ‘cite’ itself is avoided by asking the subject to instead point out the most proper method of quoting the information in a paper. Also, the excerpt provided is not difficult to understand, which was a problem with the previous versions of the question. This simplification allows subjects to focus on the actual question being asked and not on understanding what they read.

**Question 6:** For a 20th-century history class, you need to find an interview of someone’s firsthand account of a riot that happened 20 years ago in Chicago. Which of the following would provide the most appropriate results?

A. newspaper database  
B. a scientific journal  
C. Google search  
D. magazine database

The simplicity of this question is its strength; it gives a reasonable real-life academic scenario and asks the subject to distinguish possible resources based on their applicability. The subject taking the test must know what each of these resources is and what kind of information it yields in order to choose the correct answer, choice A. As usual, a tempting distractor is present in the form of Google search.
7.1.3 Item Design: A Case Study

The world has had to adapt to an incredible explosion in the availability of information in a very short time; even the mathematics to describe how the Internet works are new. Because of this, IL item construction about search technologies is on the frontier of test design. By extension, revision of such items is difficult because what exactly the cause of an item’s poor performance is can be obscured by the complex interactions of the Internet. Question three on the test suffers from these unforeseen side effects. Deconstructing the reasons behind its issues provides a useful view into what problems need to be considered when designing a good IL test item.

Question 3: Which of the following sets of keywords would be most appropriate if you were searching for information on the spread of influenza in the Middle Ages using a search engine like Google?

A. influenza, Middle Ages
B. influenza, flu, epidemic, Middle Ages, medieval
C. influenza, disease, illness, information, research, study, medieval, Middle Ages, statistics

We thought that this would be a fairly easy item, but it turned out to be an explosive example that showed just how complex and important the structure of information is in the modern age when it comes to testing a person’s understanding of it. Initially, the question seemed to be confusing for many students. Difficulties were encountered in the first pilot when one subject asked the nature of the implicit Boolean operator that the query was using. We had assumed that the obvious default search paradigm was Google, an engine which searches by keyword and relevancy, a combination which makes it very powerful without the need for explicit Boolean operators. Assuming that not specifying the querying system was a cause for confusion, we altered the question to do so. However, answers we were still not satisfied; there seemed to be something wrong with the question.

It turns out it was the subject matter and the options themselves that lead the question into trouble. In Google, Answer A actually returns results that are the most relevant, whereas answer B has a great deal of information concerning the 1918 flu epidemic. Answer C yields modern medical information about influenza.

The question fails in the real world because of how Google works. Google’s search engine ranks results by their “relevancy”, which is measured by how connected a page is to other pages, and how connected those pages are to still others.
While this has overwhelmingly proved to be a solid strategy for everyday searches, it has drawbacks. In this case, the subject matter has two looming, named historical events associated with it: the 1918 influenza epidemic and the H1N1 influenza strain. It does not take much to tip the query towards either event because of the media firestorm over H1N1. This caused the search volume to skyrocket (see Figure 7.1). If search volume surrounding a topic is considered a measure of the demand for information regarding that topic, it follows that the amount of resources providing that information will increase. Given the circumstances of the search volume spike, it can be assumed that a massive spike in information surrounding H1N1 also occurred. Since H1N1, and by association the 1918 flu and medical flu assistance, had a boost in relevancy, tipping the query in favor of anything other than Middle Ages influenza was as easy as giving more keywords that were not era-specific. Thus the properly specified query in answer B allowed the 1918 flu to dominate the results, whereas answer A did not have enough other keywords to favor the 1918 flu. Answer C demonstrates the effect even more. So many specifiers lead to the most relevant items on the Internet about flu in terms of information demand: modern medical aide.

![Figure 7.1: Google Trends data on search volume for “flu” and “influenza”][12]
In summary, it is not only important to actually make sure a practical question works as intended, but also to take into consideration how modern information works. Influenza should have been avoided as a subject matter for this question because the residual “noise” in the search system specified was too great to yield predictable results within the context of the questions itself. The following revision of this item, while it should of course be tested before being used in an actual assessment, may alleviate a number of the issues discussed in this section:

**Revised Question 3**: Which of the following sets of keywords would be most appropriate if you were searching for primary-source material on the 1918 influenza epidemic using a search engine like Google?

A. influenza, 1918
B. influenza, 1918, primary source
C. influenza, flu, 1918, aviary, interview
D. influenza, flu, statistics

### 7.2 Potential Test Items

The outcomes-to-items mapping, discussed in Chapter 4, indicated a number of outcome-standard pairs that could be assessed but which were not addressed by the final test instrument shown in Appendix E. Not every pair of outcome and standard is assessable, but those cells marked with letters in Table 4.1 (page 28) correspond to IL topics that could be assessed in a format like the current test. These potential test items are discussed below, noting both the purpose of each proposed question and the sort of tasks it might involve.

**A** This item would assess the 2.2.* range of ACRL standards, “the information literate student constructs and implements effectively-designed search strategies”, with the “obtain resources by a variety of methods” outcome. In general, such an item would ask students to perform searches and evaluate the results. For a multiple-choice test, however, it may be more appropriate to ask students to evaluate sets of results corresponding to various (provided) search queries. In this way, we may gain some understanding of the students’ ability to assess the results of “search strategies” without

\[1\] All ACRL standards quotations in this section are taken from the original ALA/ACRL standards document [1].
requiring them to complete to open-ended and often computer-aided task of constructing search strategies themselves.

b. This item would assess ACRL standard 2.3.b, “uses various classification schemes and other systems...to locate information resources within the library or to identify specific sites for physical exploration,” using the “obtain resources by a variety of methods” outcome. This sort of item is ideally suited to a practical task, for example, asking students to locate library materials. Again, if we wish to include this sort of assessment item in a multiple-choice test, we must adapt the task slightly. In this case, it may be easiest to leave the assessment of these skills to some other method, such as instructor observation during an educational program.

c. This item would assess ACRL standard 2.1.c, “investigates the scope, content, and organization of information retrieval systems,” using the “construct search strategies” outcome. Unlike the items assessing the 2.2.* range of standards, which involve the practical details of search queries, this item would need to assess students’ abilities to “construct search strategies” on a higher level. For example, an item might ask students to identify the most appropriate database or search engine to use in a particular research context.

d. This item would assess ACRL standard 2.1.c with the “gather bibliographic data and cite resources correctly” outcome. Since other items have been designed to address the mechanics of constructing a citation, this item would most likely assess students’ ability to locate bibliographic data in various information retrieval systems. However, this sort of task seems best suited to a task-based assessment, so it would require some research to see if this item could be translated to a multiple-choice test.

e. This item would assess ACRL standard 1.2.e, “differentiates between primary and secondary sources, recognizing how their use and importance vary with each discipline”, using the “read and evaluate sources critically” outcome. In this case, the critical evaluation of sources would probably focus on identifying bias in primary sources, or comparing the perspectives of primary and secondary sources discussing the same topic.

f. This item would assess ACRL standard 3.1.a, “reads the text and selects main ideas,” using the “read and evaluate sources critically” outcome. This is probably the most self-evident potential test item, since it would be nothing
more than a garden-variety reading comprehension item. However, it should be decided whether reading comprehension falls within the scope of an IL assessment of this type, and if so, what sort of passages should be used.

7.3 Administering the Test

Any test without a method of administration is useless. While both of our pilot tests were administered on paper, this method would not be feasible for assessing an entire class of first-year students.

A number of factors affected our suggested method of administration. The number of students to which the test would be administered — around 900, based on recent first-year class sizes — is the most important factor. With a population of this size, the time needed to tabulate results from a paper-and-pencil test would be prohibitive. Thus a digital deployment is the most promising. This choice, however, makes it essentially impossible to administer the assessment to an entire first-year class at once; WPI simply does not have the facilities to provide each of 900 students a computer terminal.

One possibility would be to administer the test in multiple sessions, following the model of many New Student Orientation (NSO) programs. Students could attend an “assessment session” in groups of twenty or so, allowing the use of a single computer lab. However, it seems unlikely that so many sessions could be scheduled, proctored, and incorporated into the already-packed NSO/first-year schedule.

Our solution to this problem comes from the Designs program, developed by the WPI Office of Academic Advising to assist first-year students in selecting courses for their first two terms. (See [21].) Part of this program, as described at http://www.wpi.edu/Admin/OAA/Designs3/math.html, is an electronically-administered math placement test. After taking the test, students receive feedback from the Math department suggesting which courses would be most appropriate to take. This sort of personal feedback may only be possible because students take the test individually over a period of two months. Through discussions with the Dean of Undergraduate Studies, we have established that our assessment could be included in the Designs program as well.

The math placement tests are administered within the framework provided by the WeBWorK system, which was explicitly designed for mathematical education and assessment [29]. The system’s features and interaction style are not designed for this type of assessment, and could make implementing the test. However,
having used it once already for the math assessment, students would be somewhat familiar with the interface. WPI username recognition is another benefit.

For a simpler system, especially for smaller surveys and assessments, we recommend an alternative web system, like Google Docs “Forms” service or SurveyMonkey.com, which allow users to create interactive web-based forms containing items with many formats. Gordon Library has already used Google Docs Forms for some internal assessment, so data collection with this service should not present any significant learning curve. In addition, the Google Docs platform allows multi-user access and access control by default, which would support collaboration among departments and individuals. A proof-of-concept Forms version of our assessment is shown in Appendix F.

In terms of a schedule for administering the test, we recommend a round of test revision taking our above notes into account be completed before the new school year. This revised version can then be piloted at the beginning of classes on a fresh population of first-year students. Through the fall terms, the information from this piloting should then be compiled and analyzed, leading to another revision and pilot as seen fit and in accordance with the test development process detailed in Chapter 4. This schedule allows the test to be deployed to incoming first-year students in the summer of 2011.

### 7.4 Assessment Vectors Revisited

While this project focused on a combination of a baseline skills test with first-year IL education, each of the vectors discussed in Section 3.1 remains a valid option for improving IL education and assessment in Gordon Library and on the WPI campus as a whole. The skills test is by far the most common method of IL assessment, and so any effort spent on improving and validating the test created by this project will be worthwhile.

Another option is to ask students to evaluate their own ability to use IL skills both immediately following and several weeks after an IL instruction program. This would provide information on how well students feel they remember the skills they learned and thus indicate the program’s efficacy. In a similar vein, measuring the continued use of IL skills during their academic and professional careers would provide a way to assess how useful the Library’s work was in the practical world. It would necessarily involve speaking with project groups, alumni, and employers of WPI graduates to ascertain what skills are being used in the workplace and if the Library had provided sufficient instruction.
Many useful data regarding student IL skills and learning are lost due to the lack of a method to collect them. Recording the expectations and subsequent observations of Library instructors would provide a structured record to analyze later.

Introducing students to university-level research skills in the first year to promote positive IL habits and confidence in ability to find information is a highly cited method to improve IL skills overall. With no ability to require students to participate in Library internal programs, integration with campus-wide curriculum is the best choice to improve IL education. For instance, procuring observations and hard performance data from faculty whose courses use the Library’s services would further serve the effort to reflect on how well the Library’s programs work. We would also suggest the development of a program integrated with current NSO activities, developed to be a fun, competitive standalone activity geared towards RAs and CAs and their residents. This type of integration would probably require coordination with Christine Girouard in the Student Activities Office and Kristin Murphy in Residential Services. Alternatively, a similar module could be developed for integration into the Insight program, a first-year advising and mentoring program at WPI.

7.5 Reference Guides

In light of the conclusions derived from the pilots, we also discovered a potential quick-fix solution to improving the overall level of IL skills on campus. Since our process shows that students need help on how to access research resources, and since that kind of information is concerned with “what” and “where”, a cheat-sheet of sorts is an obvious choice. The library can produce a letter-sized printout that details the precise resources students can use and where they can access them. It also would include a simple check-list as well for the best methods for refining a search query under any search paradigm, as well as the kinds of search queries. A quick reminder to ask the librarians is also highly advisable, since the ethos of re-inventing the wheel when you could have asked is not uncommon amongst engineering students. Laminated and kept at all the computers in the Library, as well as posted on the library homepage, the reference guide would be an easy addition to the Library’s arsenal.
Chapter 8

Conclusion

The field of information literacy is, to put it simply, vast. From the seven dimensions proposed by Shapiro and Hughes to the dozens of standards published by the ACRL, the information-literate individual must develop an extensive repertoire of skills and abilities. As a result, it is no simple task to assess students’ IL knowledge, not least of all because selecting a focused area to assess is almost always a significant part of the assessment process. For this project, we were able to develop a selection of possible assessment techniques that covered a variety of IL topics. We then worked with Gordon Library staff and administrators to determine which of these ‘assessment vectors’ would provide the most immediately-useful data to the Library.

One of the Library’s primary concerns, supported by much of the current IL research, is the development of IL skills among first-year students. As a result, we chose to develop a practical skills test to assess the IL knowledge and skills of first-year students upon their arrival at WPI. Because our assessment was developed specifically to assess a set of IL skills that we expect students to develop in the first year, it provides a more accurate picture of student skill level as first-year students; most extant IL assessments are designed to assess students in any class year, although of course the results of these assessments can be segmented by class year.

As discussed previously, there are a number of revisions that can be made to the assessment instrument as it stands now. These changes should help make the test a more accurate measure of IL skill level, in turn providing more useful data to the Library. Even with the additional time required to make and pilot-test these changes, it should be feasible to deploy this test on a large scale within a year (to assess the incoming class of Fall 2011 at the latest).
The IL assessment process never ends, and to that end we once again point to the original set of assessment vectors developed at the start of this project. Gordon Library should continue to develop the first-year assessment program, of course, but as resources become available, the Library should also pursue other types of assessment as well.

In the end, it is worth remembering the advice of Barclay [3]: to do the best assessment we can with what we have. Even a short assessment can prove immensely useful, whether by pointing to the need for a new program or confirming the effect of an existing one. The important thing is to continue to assess, and to evaluate the assessment program itself, making whatever adjustments are necessary to provide the most useful results for that unique situation.
Bibliography


56


URL [http://www.plagmada.org/Home.html](http://www.plagmada.org/Home.html)


   URL http://www.wpi.edu/about/policies/values.html

[31] WPI Faculty & Board of Trustees (1987). The Goal of WPI.  
   URL http://www.wpi.edu/about/mission.html

   URL http://www.wpi.edu/academics/Library/Help/Instruction/Forms/facattendeval.html

Appendix A

Sample Casual Assessment Form

**casual assessment form**

**PROGRAM:**               **AUDIENCE:**

**DATE:**                 **INSTRUCTOR:**

**DESCRIPTION:**

---

How will you focus your observation?
What behaviors or educational goals will be your focus?

---

*Remember to leave time for self-reflection after the program!*

What did you observe? Focus on the goals and topics you identified above.

---

Conclusions and Next Actions:

---

SAMPLE
Appendix B

Informed Consent Agreement for Participation in a Research Study

Investigator: “Assessing Information Literacy at WPI” IQP Group
Contact Information: libiqp@wpi.edu
Title of Research Study: “Assessing Information Literacy at WPI — Initial Pilot”
Sponsor: Gordon Library

Introduction

You are being asked to participate in a research study. Before you agree, however, you must be fully informed about the purpose of the study, the procedures to be followed, and any benefits, risks or discomfort that you may experience as a result of your participation. This form presents information about the study so that you may make a fully informed decision regarding your participation.

Purpose of the study

This study seeks to evaluate a set of test items designed to measure students’ information literacy [IL] skill level. Information literacy refers to the skill set including research skills, critical thinking, resource evaluation, etc.

Procedures to be followed

You will be asked to read and complete three or four test items relating to IL. Answer each to the best of your ability. After completing the items, a facilitator will discuss them with you, focusing on the clarity and purpose of the items.
Risks to study participants
There are no foreseeable risks or discomforts to subjects participating in this study.

Benefits to research participants and others
There are no known benefits to subjects participating in this study.

Record keeping and confidentiality
Records of your participation in this study will be held confidential so far as permitted by law. However, the study investigators, the sponsor or its designee and, under certain circumstances, the Worcester Polytechnic Institute Institutional Review Board (WPI IRB) will be able to inspect and have access to confidential data that identify you by name. Any publication or presentation of the data will not identify you.

Compensation or treatment in the event of injury
There is no foreseeable possibility of physical injury as a result of participating in this study. Nonetheless, you do not give up any of your legal rights by signing this statement.

For more information about this research or about the rights of research participants, or in case of research-related injury, contact:

IQP Team: libiqp@wpi.edu

IRB Chair (Professor Kent Rissmiller):
Tel. 508-831-5019
Email: kjr@wpi.edu

University Compliance Officer (Michael J. Curley):
Tel. 508-831-6919
Email: mjcurley@wpi.edu

Your participation in this research is voluntary. Your refusal to participate will not result in any penalty to you or any loss of benefits to which you may otherwise be entitled. You may decide to stop participating in the research at any
time without penalty or loss of other benefits. The project investigators retain the right to cancel or postpone the experimental procedures at any time they see fit.

By signing below, you acknowledge that you have been informed about and consent to be a participant in the study described above. Make sure that your questions are answered to your satisfaction before signing. You are entitled to retain a copy of this consent agreement.

____________________ Date: __________________
Study participant signature

____________________
Study participant name (please print)

____________________ Date: __________________
Signature of person who explained this study
Appendix C

WPI IRB Application and Approval

Included with the original application was a copy of the first pilot test instrument, as shown in Appendix D. More information about the WPI IRB approval process can be found at http://www.wpi.edu/Admin/IRB/index.html.
Use of this application is recommended for most student project research involving minimal risk. Proposed research meets the definition of “minimal risk” when the risks to research subjects are not greater than those ordinarily encountered in daily life. This application is specifically intended for projects in which students are expected to conduct interviews, surveys or focus groups. Please return a signed hard or electronic copy of this application to the WPI IRB c/o Ruth McKeough, 2nd floor Project Center or irb@wpi.edu. If you have any questions, please call (508) 831-6699.

Project Faculty Advisor(s):

Name: Arthur Heinricher
Department: Dean of Undergraduate Studies/Mathematical Sciences
Tel No: E-Mail Address:

Name: Christine Drew
Department: Gordon Library
Tel No: E-Mail Address:

Student Investigator(s):

Name: Ben Laverriere
Name: Michael Olver
Name: Madison Dickson
Tel No: E-Mail Address:
Tel No: E-Mail Address:
Tel No: E-Mail Address:

Project Title: Assessing Information Literacy At WPI [IGP]

Project Location and Time Frame: WPI Campus, through C Term 2010

Expected Research Subjects: (e.g. museum visitors under the age of 12)
WPI students, first- and second-year.

NOTE: This application must be accompanied by written research methods and a reasonably complete set of survey or interview questions.

1. Is the proposed research sponsored or supported by a US federal agency or by US government funding?

   No ☒ Yes ☐

2. Is the proposed research funded by a corporation or foundation? If so, please identify sources.

   No ☒ Yes ☐
WPI IRB Application for Exemption from IRB Review for 
Survey or Interview Research Involving Minimal or No Risk

3. Does the proposed research involve vulnerable research subjects? (e.g. children, prisoners, students, persons with mental or physical disabilities, pregnant women)
   - No [X] Yes [O]

4. Is the research confined to obtaining verbal or written information from subjects and/or publicly available documentary information?
   - No [O] Yes [X]

5. Could the disclosure of a human subject’s identity and responses place the subject at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation?
   - No [X] Yes [O]

6. Will the researchers collect information that can be used to identify the subjects?
   - No [X] Yes [O]

7. If the researchers do know the subjects’ identity, will individual responses be kept confidential? (e.g. only summaries of all data will be published)
   - No [O] Yes [O]

8. Will researchers be interviewing people chosen because of their expertise or experience? (See 4, below.)
   - No [X] Yes [O]

Please Print Form before signing below

By signing below, all participants in this research project are agreeing to follow the following instructions:

1. You agree to inform subjects orally or in writing that:
   - Participation in the research is voluntary.
   - Participants may end their participation at any time.
   - Participants need not answer every question in an interview or survey.

2. If your research is anonymous, you also inform subjects that you are not collecting names or any identifying information from them.

3. If your research is confidential, you inform subjects that no identifying information will be disclosed with individual responses.

4. If your research subjects are chosen and interviewed for their expertise or experience, you seek and obtain each subject’s permission to identify him or her in your report, and obtain each subject’s permission to disclose his or her views and statements in your report. The subject must be offered the opportunity to pre-approve the publication of any quoted material. If a subject does not wish to appear in your report, you respect his or her wishes for confidentiality.

Signature of Faculty Advisor
[Signature]
Date 12/4/09

Print Full Name and Title
Arthur C. Heinricher, Dean of Undergraduate Studies
Assessing Information Literacy at WPI - Proposed Research Methods

This project seeks to develop an assessment instrument for use at Gordon Library to assess the information literacy skills of first-year students at WPI. (Information literacy, as defined by the Association of College and Research Libraries, refers to "a set of abilities requiring individuals to recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information."*) An important facet of test development, as indicated in the attached flowchart, is the cycle of pilot testing and revision, to ensure that the instrument can accurately measure the desired data. As such, the research for this project — or rather, research involving human subjects — falls into two phases. The first phase, pending approval by the IRB, will be completed December 9 through 12, 2009; the second phase is currently unscheduled but will occur during or before C term 2010. A brief description of each phase follows, including an overview of the method of administration and data collection.

**Phase One: Small-Scale Pilot Tests**

The primary purpose of this phase is to ensure that the current versions of the test items are, first, comprehensible, and that they address the particular skills they are attended to assess. To that end, we hope to offer a series of pilot sessions, aiming for a subject population of thirty students or more, wherein subjects would be given a set of four draft items to complete. (In addition, demographic data will be collected from each student [class year, amount of library instruction] but no personally identifying information will be collected.) After completing the items, the subjects will be presented with alternate forms of those items (with regard to phrasing, answer format, etc.) and will discuss with a facilitator whether any of the alternate forms seem more understandable to the subject. The subjects' responses in this portion of the study will be recorded in the form of anonymous notes by the facilitators, to be used in the next cycle of revision.

**Phase Two: Full-Scale Administration**

At this point, the researchers will have constructed a full assessment, based on the results of Phase One. Pending collaboration with administrators, we hope to administer the instrument to incoming first-year students (transfer students and late-admission students) towards the beginning of C term. The administration of the test, by design, will take approximately 20 minutes. Similar demographic data to Phase One will be collected, and the subjects will be asked to complete the full set of items. Again, no personally-identifying data will be collected.

Re: IRB Application for Exemption 2009-EX-114 “Assessing Information Literacy at WPI - IQP”

Dear Professor Heinricher,

The WPI Institutional Review Committee (IRB) has reviewed the materials submitted in regards to the above mentioned study and has determined that this research is exempt from further IRB review and supervision under 45 CFR 46.101(b)(2): “Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and any disclosure of the human subjects’ responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects’ financial standing, employability, or reputation.”

This exemption covers any research and data collected under your protocol from 7 December 2009 to 6 December 2010, unless terminated sooner (in writing) by yourself or the WPI IRB. Amendments or changes to the research that might alter this specific exemption must be submitted to the WPI IRB for review and may require a full IRB application in order for the research to continue.

Please contact the undersigned if you have any questions about the terms of this exemption.

Thank you for your cooperation with the WPI IRB.

Sincerely,

[Signature]

Kent Rissmiller
WPI IRB Chair
Appendix D

Pilot 1 Test Instrument

On the following pages are reproduced the demographic survey and test instrument used in Pilot 1, followed by a chart of the data collected. This test was piloted to 18 students during the following times:

- Wednesday, December 9th, 2009: 1pm-3pm, Kaven Hall 202
- Thursday, December 10th, 2009: 11am-1pm, Kaven Hall 202
- Saturday, December 12th, 2009: 2pm-4pm, Gordon Library (2F)
Information Literacy Survey

What is your class year?    _____ Freshman    _____ Sophomore

What is your intended Major? (Use “undecided” if still undecided) ____________________

On average, how many hours a week do you spend in the library?

_____ A. I hardly ever go to the library
_____ B. 1-3 hours
_____ C. 4-7 hours
_____ D. 8+ hours

In the past two terms, have you taken a course that required you to do research for a paper or project?

_____ Yes    _____ No

Which of the following Library resource instructions have you used or received? (check all that apply)

☐ Class presentation
☐ One on One or Group Consultation
☐ Online Text Tutorial
☐ Online Video Tutorial

Answers:
Test Type: ________________  (Written on top of questions page)

1. _______

2. _______

3. _______

4. _______
**Question 1**
You are researching the spread of influenza in the Middle Ages. Which set of keywords will give you the most relevant results?

A. influenza, Middle Ages  
B. influenza, flu, epidemic, Middle Ages, medieval  
C. influenza, disease, illness, information, research, study, medieval, Middle Ages, statistics

**Question 2**
"Semiotics can be conceived of either as a unified theoretical approach to the great variety of systems of signification and communication, and in this sense it constitutes a metalinguistic discourse dealing with any of its objects by means of homogeneous categories, or it can be conceived as a description of those various systems insisting on their mutual differences, their specific structural properties, their idiosyncrasies — from verbal language to gestures, from visual images to body positions, from musical sounds to fashions. It shows a wide range of "languages" ruled by different conventions and laws. It can investigate those various domains either at the elementary level of their consecutive units (such as words, color spots, physical formants of sounds, geometrical or topological shapes) or at the more complex level of texts and discourses — that is, narrative structures, figures of speech and so on."


If you were writing a paper, which of the following is the appropriate way to incorporate the above ideas into your writing?

A. Semiotics can be conceived of either as a unified theoretical approach to the great variety of systems of communication, or as a description of those various systems insisting on their mutual differences. It can investigate those various domains either at the elementary level of their consecutive units (such as words, color spots, physical formants of sounds, geometrical or topological shapes) or at the more complex level of texts and discourses — that is, narrative structures, figures of speech and so on.

B. Umberto Eco defined semiotics as "a unified theoretical approach to the great variety of systems of signification and communication". He suggests that there is a wide range of "languages" that contribute to communication, both at the elementary level (words, color spots, etc.) and at the more complex level of texts and discourses.

C. The field of semiotics explores the ways in which forms of communication are given meaning by human beings. Umberto Eco suggests that this investigation can occur "either at the elementary level of their consecutive units (such as words, color spots, physical formants of sounds, geometrical or topological shapes) or at the more complex level of texts and discourses — that is, narrative structures, figures of speech and so on" (Eco, 108).

**Question 3**
Who can access the Library's information databases?

A. students and faculty  
B. faculty and alumni  
C. students and alumni  
D. students
Question 1
A truncation search for "nanotech" in the library catalog will include results related to which set of terms?

A. nano, nanos, nanosized, nanobit
B. nanotechnology, nanotechnological, nanotechnologies
C. nanostructured, nanoparticles, nanophotonics
D. nano-micro, nanoscale, nanocrystalline

Question 2
For a 20th century history class, need to find a firsthand account of a riot that happened 20 years ago in Chicago. Which of the following would provide the most appropriate results?

A. newspaper database
B. a journal
C. Google search
D. magazine database

Question 3
Which of the following is a valid citation style?


Question 4
You need to look up the beliefs of an ancient religion called Bardism. Which of the following sources seems more reliable?

A. "Amazing World!" a section on cultural subjects from a local newspaper
B. "Bardism and You" from website www.thebardlives.com
C. "Chariots of the Bards" from website www.bard-spaceships.com
D. "The Way of the Bard" a book by Professor Dick Smith
**Question 1**
For your plant biology class, you must research a species of plant. A place you'd expect to find good results would be:

A. scientific journal  
B. a Google search  
C. magazine  
D. newspaper database

**Question 2**
"Gadgets" are widely used but controversial devices. You are researching their environmental impact for a paper. Which of the following seems to be the least biased source?

A. "Pollution and Gadgets" by the National Gadget Makers Association  
B. "What the Government Won't Tell You About Gadgets" from the news blogging site Web Voice  
C. "Building a Better Gadget" an editorial in the New York Times  
D. "Gadgets and Regional Pollution" from the Journal of Gadgets

**Question 3**
You've been assigned a research paper on the effects of acid rain on the fishing industry in Canada. If you're interested in finding documents related to this topic that specifically include the phrase "acid rain" or the phrase "acid precipitation", which of the following Boolean searches would work best?

A. fisheries AND Canada AND (acid rain OR acid precipitation)  
B. fisheries AND acid rain OR acid precipitation AND Canada  
C. fisheries OR Canada OR (acid rain OR acid precipitation)  
D. fisheries AND (acid rain AND acid precipitation) AND Canada

**Question 4**
Which of the following is not a valid citation style?

**Question 1**
You need to look up the beliefs of a religion called Neo-Paganism.  
Which of the following sources would be most credible?

A. "Neo-Paganism" from the Wikipedia article at www.wikipedia.org/wiki/Neo-Paganism  
B. "The Pagan Diaries" an editorial from the local newspaper  
C. "Neo-Paganism: from Humble Beginnings" an article from the journal Philosophy and Religion Quarterly  
D. "The Truth about Paganism" from the website www.thesecrettruth.net/revealed.html

**Question 2**
You've been assigned a research paper on the effects of acid rain on the fishing industry in Canada. To retrieve documents/results with EITHER of the following phrases present:
-“Acid rain”
-“Acid precipitation”
Which search below would work best?

A. fisheries AND Canada AND (acid rain OR acid precipitation)  
B. fisheries AND acid rain OR acid precipitation AND Canada  
C. fisheries OR Canada OR (acid rain OR acid precipitation)  
D. fisheries AND (acid rain AND acid precipitation) AND Canada

**Question 3**
Who can access the Library's information databases?

A. students and faculty  
B. faculty and alumni  
C. students and alumni  
D. students

**Question 4**
Which of the following sets of keywords would be most appropriate if you were searching for information on the spread of influenza in the Middle Ages?

A. influenza, Middle Ages  
B. influenza, flu, epidemic, Middle Ages, medieval  
C. influenza, disease, illness, information, research, study, medieval, Middle Ages, statistics
## D.1 Pilot 1 Results

<table>
<thead>
<tr>
<th>Year</th>
<th>Major</th>
<th>Time in Library</th>
<th>Coursework</th>
<th>Resource instruction</th>
<th>Test Type</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>PH</td>
<td>a</td>
<td>y</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>iguana</td>
</tr>
<tr>
<td>2</td>
<td>ME</td>
<td>a</td>
<td>y</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>iguana</td>
</tr>
<tr>
<td>2</td>
<td>CHE</td>
<td>d</td>
<td>y</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>iguana</td>
</tr>
<tr>
<td>2</td>
<td>CHE</td>
<td>d</td>
<td>y</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>iguana</td>
</tr>
<tr>
<td>2</td>
<td>CS</td>
<td>b</td>
<td>n</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>iguana</td>
</tr>
<tr>
<td>2</td>
<td>CHE</td>
<td>d</td>
<td>y</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>iguana</td>
</tr>
<tr>
<td>2</td>
<td>CHE</td>
<td>d</td>
<td>y</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>iguana</td>
</tr>
<tr>
<td>2</td>
<td>Psy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Science</td>
<td>b</td>
<td>y</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>ostrich</td>
</tr>
<tr>
<td>2</td>
<td>Act. Math</td>
<td>c</td>
<td>y</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>ostrich</td>
</tr>
<tr>
<td>2</td>
<td>CHE</td>
<td>c</td>
<td>y</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>ostrich</td>
</tr>
<tr>
<td>2</td>
<td>CS</td>
<td>a</td>
<td>y</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>ostrich</td>
</tr>
<tr>
<td>2</td>
<td>CHE</td>
<td>d</td>
<td>y</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>ostrich</td>
</tr>
<tr>
<td>2</td>
<td>ECE</td>
<td>c</td>
<td>n</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>ostrich</td>
</tr>
<tr>
<td>2</td>
<td>Indus Eng</td>
<td>d++</td>
<td>y</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>ostrich</td>
</tr>
<tr>
<td>2</td>
<td>Indus Eng</td>
<td>b</td>
<td>y</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>zebra</td>
</tr>
<tr>
<td>2</td>
<td>CHE</td>
<td>c</td>
<td>y</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>Zebra</td>
</tr>
<tr>
<td>2</td>
<td>Civil &amp; FPE</td>
<td>c</td>
<td>y</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>Zebra</td>
</tr>
<tr>
<td>2</td>
<td>FPE</td>
<td>b</td>
<td>n</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>zebra</td>
</tr>
<tr>
<td>1</td>
<td>CS</td>
<td>c</td>
<td>n</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>zebra</td>
</tr>
</tbody>
</table>

Correct answers are highlighted in green.

Answers: Iguana C A A B
Ostrich B A A D
Penguin A C A
Zebra A D A C
Appendix E

Pilot 2 Test Instrument

On the following pages are reproduced the pilot test instruments used in Pilot 2, followed by a chart of the data collected. Correct answers are highlighted in green. This test was piloted to 17 late-admit students during the NSO Library information session at 9AM on Tuesday, January 12th, 2010 in Anderson Lab B.
Information Literacy Assessment

This pilot test is designed to assess the information literacy skill level of incoming first-year students at Worcester Polytechnic Institute. It has been produced by an IQP team working with Gordon Library Instruction and Outreach staff, with the purpose of helping educate WPI students about research techniques and critical thinking skills.

This test will not affect the grade of any student who takes it. It is purely a diagnostic tool to allow the Gordon Library Instruction and Outreach staff to target student information literacy needs.

The test has two sections: a short data collection survey and a section of multiple choice questions. It should take around 12 minutes. Circle your answers directly on the test.

Survey

What is your intended major? (use “undecided” if you have not declared a major)

________________________________________

Which status best describes you?

□ a first-year student with no prior college experience
□ a first-year student with prior college experience (and not a transfer student)
□ a transfer student (not a first-year student)

Which of the following library research education experiences have you used or participated in previously? (check all that apply)

□ a class presentation
□ a one-on-one or group consultation with a librarian
□ an online text tutorial
□ an online video tutorial

Questions

(1) Research scenario: You need to research the beliefs of a religion called Neo-Paganism. Which of the following sources would be most credible for a detailed research project?

A. "Neo-Paganism", from the Wikipedia article at www.wikipedia.org/wiki/Neo-Paganism

B. "The Pagan Diaries", an editorial from the local newspaper

C. "Neo-Paganism: from Humble Beginnings", an article from the journal Philosophy and Religion Quarterly

D. "The Truth about Paganism", from the website www.thesecrettruth.net/revealed.html

(2) Who can access the Library's information databases?

A. students and faculty
B. students, faculty, and alumni
C. students and alumni
D. students only

(3) Which of the following sets of keywords would be most appropriate if you were searching for information on the spread of influenza in the Middle Ages using a search engine like Google?

A. influenza, Middle Ages
B. influenza, flu, epidemic, Middle Ages, medieval
C. influenza, disease, illness, information, research, study, medieval, Middle Ages, statistics

(4) Read the following text and answer the question at the end.

"Worldwide during 1995, 2.7 million new adult HIV infections occurred (averaging more than 7,000 new infections each day). Of these, about 1 million (an average of nearly 3,000 new infections per day) occurred in Southeast Asia, and 1.4 million were in sub-Saharan Africa. The industrialized world accounted for about 55,000 new HIV infections in 1995 (nearly 150 new infections per day: about 2 percent of the global total)."


If you were writing a paper about the spread of HIV/AIDS in third-world countries, which of the following is the most appropriate way to quote the author's ideas above into your writing?

A. In 1995, 2.7 million new adult HIV infections occurred (averaging more than 7,000 new infections each day). Of these, about 1 million (an average of nearly 3,000 new infections per day) occurred in Southeast Asia, and 1.4 million were in sub-Saharan Africa.

B. Chris Beyrer writes that in 1995, "2.7 million new adult HIV infections occurred", meaning there were more than 7,000 new infections each day. Of these, "about 1 million" occurred in Southeast Asia, and 1.4 million were in sub-Saharan Africa.

C. In 1995, new adult HIV infections averaged "more than 7,000 new infections each day", with about 3,000 of these occurring in Southeast Asia. Further, the unindustrialized world accounted for about 98 percent of new HIV infections in that year (Beyrer, 3).

(5) A truncation search for "nanotech" in the library catalog will include results related to which set of terms?

A. nano, nanos, nanosized, nanobit
B. nanotechnology, nanotechnological, nanotechnologies
C. nanostructured, nanoparticles, nanophotonics
D. nano-micro, nanoscale, nanocrystalline

(6) For a 20th-century history class, you need to find an interview of someone's firsthand account of a riot that happened 20 years ago in Chicago. Which of the following would provide the most appropriate results?
(7) For your plant biology class, you must research a species of plant. A place you'd expect to find relevant and reliable information would be:

A. a peer-reviewed journal
B. a Google search
C. a magazine
D. a newspaper database

(8) "Gadgets" are widely used but controversial devices. You are researching their environmental impact for a paper. Which of the following seems to be the most likely source for objective information?

A. "Pollution and Gadgets" by the National Gadget Makers Association
B. "What the Government Won't Tell You About Gadgets" from the news blogging site Web Voice
C. "Building a Better Gadget" an editorial in the New York Times
D. "Gadgets and Regional Pollution" from the Journal of Gadgets

(9) You've been assigned a research paper on the effects of acid rain on the fishing industry in Canada. To retrieve results with EITHER of the phrases "acid rain" or "acid precipitation", which query below would work best?

A. fisheries AND Canada AND (acid rain OR acid precipitation)
B. fisheries AND acid rain OR acid precipitation AND Canada
C. fisheries OR Canada OR (acid rain OR acid precipitation)
D. fisheries AND (acid rain AND acid precipitation) AND Canada

(10) You are writing a bibliography. Which of the following is not a valid citation for a journal article?

### E.1 Pilot 2 Results

<table>
<thead>
<tr>
<th>Major</th>
<th>Status</th>
<th>Previous</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1</td>
<td>1000</td>
<td>C</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Physics</td>
<td>2</td>
<td>0010</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>D</td>
<td>B</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>Robotics</td>
<td>1</td>
<td>1000</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>D</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>ECE</td>
<td>1</td>
<td>1000</td>
<td>D</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>D</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>undecided</td>
<td>2</td>
<td>1001</td>
<td>C</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>D</td>
<td>A</td>
<td>D</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Robotics</td>
<td>2</td>
<td>0011</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>ME</td>
<td>2</td>
<td>0010</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>D</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>3</td>
<td>1110</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>ME</td>
<td>2</td>
<td>1111</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Aerospace</td>
<td>1</td>
<td>1000</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>4</td>
<td>1111</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>2</td>
<td>1000</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>D</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>CS</td>
<td>2</td>
<td>1010</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>C</td>
<td>D</td>
<td>A</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>ME</td>
<td>2</td>
<td>1110</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>Civil</td>
<td>1</td>
<td>1010</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>D</td>
<td>B</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>ECE</td>
<td>3</td>
<td>0001</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>D</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>undecided Engineer</td>
<td>2</td>
<td>1011</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>

**Answers:** C A B C B A A D A C

**Num Correct:** 15 7 5 10 15 14 8 9 9 10

**Status Key:**
- 1: first year w/ no prior experience
- 2: first year w/ prior experience
- 3: transfer student
- 4: exchange student

**Previous Key:**
- 1: a class presentation
- 1___ : one on one or group consult with a librarian
- ___1 : online text tutorial
- ___1 : online web tutorial
Appendix F

Web Administration Mockup

Following is an example of how an IL assessment could be deployed using Google Docs Forms. The service allows users to download the resulting data in a variety of formats, such as the Excel spreadsheet shown below.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Timestamp</td>
<td>Survey 1</td>
<td>Survey 2</td>
<td>Question 1</td>
</tr>
<tr>
<td>2</td>
<td>2/11/2010 8:36:09</td>
<td>Humanities and</td>
<td>a first-year student</td>
<td>&quot;Neo-Paganism:</td>
</tr>
<tr>
<td>3</td>
<td>2/11/2010 12:33:15</td>
<td>Computer Science</td>
<td>a first-year student</td>
<td>&quot;Neo-Paganism:</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The address to view this page online is https://spreadsheets.google.com/viewform?hl=en\&formkey=dD1DNUNoNjU0SmRSZm5aUWtDV09aeEE6MA
A screenshot of the online form is included on the following page.
Prototype IL Assessment

This pilot test is designed to assess the information literacy skill level of incoming first-year students at Worcester Polytechnic Institute. It has been produced by an IQP team working with Gordon Library Instruction and Outreach staff, with the purpose of helping educate WPI students about research techniques and critical thinking skills.

This test will not affect the grade of any student who takes it. It is purely a diagnostic tool to allow the Gordon Library Instruction and Outreach staff to target student information literacy needs.

The test has two sections: a short data collection survey and a section of multiple choice questions. It should take around 12 minutes.

* Required

Survey 1 *
What is your intended major? (use “undecided” if you have not declared a major)

Survey 2 *
Which status best describes you?
- a first-year student with no prior college experience
- a first-year student with prior college experience (and not a transfer student)
- a transfer student (not a first-year student)
- a foreign exchange student (of any class year)

Question 1 *
Research scenario: You need to research the beliefs of a religion called Neo-Paganism. Which of the following sources would be most credible for a detailed research project?
- "Neo-Paganism", from the Wikipedia article at www.wikipedia.org/wiki/Neo-Paganism
- "The Pagan Diaries", an editorial from the local newspaper
- "Neo-Paganism: from Humble Beginnings", an article from the journal Philosophy and Religion Quarterly
- "The Truth about Paganism", from the website www.thesecrettruth.net/revealed.html

Submit
Appendix G

Test Specification Document

G.1 Purpose

The test instrument described in this document has been designed to assess the information literacy (IL) skills of first-year students, using a subset of the Association of College and Research Libraries (ACRL) standards for IL competency. Specifically, the test was created to be a baseline skills assessment, to be performed shortly before or upon students’ arrival at Worcester Polytechnic Institute (WPI). However, we believe that the test could also be used as a post-test at the end of the first year, since students are expected to have developed the skills being assessed by the test in their first year.

The development process used to create this test is based largely on the work of the Bay Area Community Colleges Information Competency Assessment Project (ICAP) [25], and similarly this document is based on the specification document created for ICAP [4]. Further details about the construction of this test can be found in Dickson et al. [8].

G.2 Learning Outcomes and Standards

One of the guiding principles of this test’s development was that it should have a clear and explicit connection to existing IL standards. To that end, the following set of ACRL IL standards were chosen as “assessable standards” to be addressed by the test:

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

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

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

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

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

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

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

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

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

3.1.a Reads the text and selects main ideas.

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

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

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

5.1.d Demonstrates an understanding of intellectual property, copyright, and fair use of copyrighted material.
In addition, a set of practical IL learning outcomes was developed, so that test items measuring students’ ability to achieve one of the outcomes would show compliance with one or more of the ACRL standards.

A Obtain resources by a variety of methods
B Construct search strategies
C Evaluate source type, relevance, and validity
D Gather bibliographic data and cite resources correctly
E Understand and avoid plagiarism

Finally, items were constructed based on outcome-standard pairs. The mapping of test items to outcomes and standards is shown in Table G.1 below.

<table>
<thead>
<tr>
<th></th>
<th>1.2.e</th>
<th>2.1.c</th>
<th>2.2.*</th>
<th>2.5.c</th>
<th>3.2.a</th>
<th>5.1.d</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td>3, 5, 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>6, 8</td>
<td></td>
<td>1, 6, 7, 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td>4, 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

Table G.1: Students’ ability to perform the task listed in each row should correspond to mastery of the ACRL standard listed in the column header. Thus, an item number in a cell indicates an assessment item designed to test the corresponding task outcome.

G.3 Test-Level Design Decisions

Format. We chose a multiple-choice format for a number of reasons. First, constructing a scoring rubric for a multiple-choice test is far simpler than developing such a rubric for a short-answer or task-based assessment. Since our lack of experience was likely to introduce a great deal of imprecision into our test items, we chose to construct items of a form that would provide the least ambiguous option for scoring.
Further, we knew that Gordon Library, the eventual administrator of this test, had limited staff and time to devote to assessment. Following the advice of Donald Barclay [3], to “do the best assessment you can with what you have,” we thought it most prudent to make a test that could be administered and scored easily, rather than a longer, more detailed test that the Library would be unable to use.

**Timing.** We chose to make this a relatively short test for two reasons. First, the set of skills that were both assessable in a multiple-choice test and also reasonable to expect first-year students to develop was a small one, particularly when compared to the complete set of ACRL standards. Second, if this test is to be deployed to students near the beginning of their first year, we knew that Gordon Library would be competing to secure time in the students’ already-busy schedules. As a result, we set our benchmark for test administration time at twenty minutes.

**Administration.** Some IL assessments ask students to use various information resources to complete tasks as a test of various IL skills. The Interdisciplinary Research (ID2050) test used by Gordon Library is one such test. However, the format of our test does not require such tasks, and as such requires no access to any external resources. Students need not — and should not — use web sites, reference books, or any other materials when completing the assessment.

### G.4 Test Item Design Decisions

**Question 1**

Research scenario: You need to do research the beliefs of a religion called Neo-Paganism. Which of the following sources would be most credible for a detailed research project?

A. ”Neo-Paganism”, from the Wikipedia article at [www.wikipedia.org/wiki/Neo-Paganism](http://www.wikipedia.org/wiki/Neo-Paganism)

B. ”The Pagan Diaries”, an editorial from the local newspaper

C. ”Neo-Paganism: from Humble Beginnings”, an article from the journal *Philosophy and Religion Quarterly*

D. ”The Truth about Paganism”, from the website [www.thesecrettruth.net/revealed.html](http://www.thesecrettruth.net/revealed.html)

Question 1 focuses on web-based sources and their trustworthiness. The correct answer is ‘C’, as the peer-reviewed nature of journals makes Philosophy and
Religion Quarterly a very likely source for solid information. Answer ‘A’ with Wikipedia serves as a distractor, as the site often serves as a good source of general information and usually is a good starting place. The item corresponds to the outcome scenario of evaluating resources, creating search strategies, and deals with standards 2.1.c, 3.2.a, and 2.3.b.

**Question 2**: Who can access the Library’s information databases?

A. students and faculty  
B. students, faculty, and alumni  
C. students and alumni  
D. students only

Question 2 is designed to test students’ knowledge of a particular policy at Gordon Library. The correct answer is ‘A’. It should be noted that this question does not in fact, align with any of the standards that we intended to assess, as it is institution-specific. While it is true that students should be aware of Library policy, we also realize that most first-year students will not be aware of this policy when they first come to WPI. Furthermore, though Gordon Library does have an official policy on the use of its databases by alumni that prohibits such use, it is a poorly communicated and sometimes insufficiently enforced one. As discussed in Dickson et al. [8], for these reasons and more, we strongly recommend revising or removing this question.

**Question 3**: Which of the following sets of keywords would be most appropriate if you were searching for information on the spread of influenza in the Middle Ages using a search engine like Google?

A. influenza, Middle Ages  
B. influenza, flu, epidemic, Middle Ages, medieval  
C. influenza, disease, illness, information, research, study, medieval, Middle ages, statistics

Question 3 tests students’ ability to evaluate various sets of keyword searches for effectiveness. The correct answer is choice ‘B’ because it uses appropriate synonyms and doesn’t add irrelevant terms like those in choice ‘C’. Because this item tests a type of search strategy, it assesses the 2.2.* family of standards.
Question 4

Read the following text and answer the question at the end.

“Worldwide during 1995, 2.7 million new adult HIV infections occurred (averaging more than 7,000 new infections each day). Of these, about 1 million (an average of nearly 3,000 new infections per day) occurred in Southeast Asia, and 1.4 million were in sub-Saharan Africa. The industrialized world accounted for about 55,000 new HIV infections in 1995 (nearly 150 new infections per day: about 2 percent of the global total).”


If you were writing a paper about the spread of HIV/AIDS in third-world countries, which of the following is the most appropriate way to quote the author’s ideas above into your writing?

A. In 1995, 2.7 million new adult HIV infections occurred (averaging more than 7,000 new infections each day). Of these, about 1 million (an average of nearly 3,000 new infections per day) occurred in Southeast Asia, and 1.4 million were in sub-Saharan Africa.

B. Chris Beyrer writes that in 1995, “2.7 million new adult HIV infections occurred”, meaning there were more than 7,000 new infections each day. Of these, ”about 1 million” occurred in Southeast Asia, and 1.4 million were in sub-Saharan Africa.

C. In 1995, new adult HIV infections averaged “more than 7,000 new infections each day,” with about 3,000 of these occurring in Southeast Asia. Further, the unindustrialized world accounted for about 98 percent of new HIV infections in that year (Beyrer, 3).

Question 4 is designed to assess the subject’s understanding of proper citation of sources. The correct answer is ‘C’, since it is the only answer that provides a reference to the actual source (with a parenthetical citation). Answer ‘A’ should be a clear example of plagiarism to any student who is familiar with the concept, while answer ‘B’ serves to enforce the importance of a proper citation rather than a vague reference to an author. This item corresponds to the “understand and avoid plagiarism” outcome, and therefore helps to assess standard 5.1.d.
Question 5

A truncation search for “nanotech” in the library catalog will include results related to which set of terms?

A. nano, nanos, nanosized, nanobit
B. nanotechnology, nanotechnological, nanotechnologies
C. nanostructured, nanoparticles, nanophotonics
D. nano-micro, nanoscale, nanocrystalline

This item was originally featured in Gordon Library’s ID2050 research quiz, and only the phrasing of the question was modified before its inclusion in our instrument. The item is intended to test the subject’s knowledge of truncation search, since only one of the possible answers (answer B) contains terms with “nanotech” as a prefix. Ideally, this item would assess the “construct search strategies” outcome, specifically standard 2.2.d. However, we have since realized that the three distractor choices (A, C, and D) are all incorrect in the same manner — they contain terms beginning with “nano” — and therefore make this question a test of pattern identification rather than truncation searching. The purpose and format of the question remain valid; the only revision necessary is the creation of a more diverse set of distractors,

Question 6: For a 20th-century history class, you need to find an interview of someone’s firsthand account of a riot that happened 20 years ago in Chicago. Which of the following would provide the most appropriate results?

A. newspaper database
B. a scientific journal
C. Google search
D. magazine database

Question 6 gives a reasonable real-life academic scenario and asks the subject to distinguish possible resources based on their applicability. This item is designed to assess the “construct search strategies” outcome and the 2.2.* family of standards. The subject taking the test must know what each of these resources is and what kind of information it yields in order to choose the correct answer, choice ‘A’. As usual, a tempting distractor is present in the form of Google search.
**Question 7**

For your plant biology class, you must research a species of plant. A place you’d expect to find relevant and reliable information would be:

A. a peer-reviewed journal  
B. a Google search  
C. a magazine  
D. a newspaper database

Question 7 assesses the subject’s knowledge of different general information resources. The correct answer is ‘A’, a peer-reviewed scientific journal. A Google search serves as a distractor in the same manner as Wikipedia does in Question 1 (see G.4). The item targets outcome scenarios evaluating sources and evaluating source type and relevancy. Standards 2.1.c, 3.2.a, and 2.3.b apply.

**Question 8**

“Gadgets” are widely used but controversial devices. You are researching their environmental impact for a paper. Which of the following seems to be the most likely source for objective information?

A. “Pollution and Gadgets” by the National Gadget Makers Association  
B. “What the Government Won’t Tell You About Gadgets” from the news blogging site Web Voice  
D. “Gadgets and Regional Pollution” from the Journal of Gadgets

In this item, an invented device called a ‘gadget’ serves as an impartial manner of dealing with subject matters that carry a great deal of controversy, legitimate or not. The test taker is asked to identify a source for information that is likely to yield unbiased information. The correct answer is ‘D’, a relevant journal. ‘A’ serves as a distractor as it sounds like a similarly ‘prestigious’ publication as the journal in ‘D’. Resources are evaluated for relevance and validity, while targeting standards 3.2.a and 2.1.c.
**Question 9**
You’ve been assigned a research paper on the effects of acid rain on the fishing industry in Canada. To retrieve results with *either* of the phrases “acid rain” or “acid precipitation,” which query below would work best?

A. fisheries AND Canada AND (acid rain OR acid precipitation)
B. fisheries AND acid rain OR acid precipitation AND Canada
C. fisheries OR Canada OR (acid rain OR acid precipitation)
D. fisheries AND (acid rain AND acid precipitation) AND Canada

Question 9 is another search-strategies question, asking students to demonstrate an understanding of Boolean search operators. This question was also adapted from the existing ID2050 test. The correct answer is ‘A’, since it requires the terms ‘fisheries’ and ‘Canada’ and either of the phrases ‘acid rain’ or ‘acid precipitation’. Like the other search-strategies questions, this item tests the 2.2.* family of standards.

**Question 10**
You are writing a bibliography. Which of the following is not a valid citation for a journal article?


Question 10 is intended to perform a baseline screening to determine if students have any understanding of the purpose and content of a bibliographic citation. We did not wish to evaluate whether subjects were familiar with any particular citation style, nor to penalize them if they were unfamiliar with some “proper” style used in the item. As a result, we chose the negative phrasing shown above, despite Radcliff’s advice (based on the findings of various test designers and researchers) to avoid this sort of construction [22]. The correct answer is choice ‘C’, since the bibliographic data provided are not sufficient for any common citation style. This item corresponds to the “gather bibliographic data and cite resources correctly” outcome, and references standard 2.5.c in particular.