Communications in Information Literacy

Volume 3 | Issue 2 Article 6

3-16-2010

Using Cognitive Interviews and Student Response Processes to Validate an Interpretive Argument for the ETS iSkills™ Assessment

Eric Snow
Center for Technology in Learning, SRI International, eric.snow@sri.com

Irvin Katz
Educational Testing Service, irkatz@yahoo.com

Recommended Citation

Snow, E., & Katz, I. (2010). Using Cognitive Interviews and Student Response Processes to Validate an Interpretive Argument for the ETS iSkills™ Assessment. *Communications in Information Literacy, 3* (2), 99-127. https://doi.org/10.15760/comminfolit.2010.3.2.75

Volume 3, Issue 2, 2009

Communications in INFORMATION LITERACY

FOCUS ON ASSESSMENT [ARTICLE]

USING COGNITIVE INTERVIEWS TO VALIDATE AN INTERPRETIVE ARGUMENT FOR THE ETS *iSKILLS*™ ASSESSMENT

Eric Snow
Center for Technology in Learning, SRI International

Irvin R. Katz
Educational Testing Service

ABSTRACT

Evaluating the trustworthiness of Internet-based or other digital information has become an essential 21st century skill. The *iSkills*™ assessment, from Educational Testing Service (ETS), purports to measure such digital evaluation skills, along with other digital literacy skills. In this work, we use an argument-based approach to assessment validation to investigate the extent to which *iSkills* test scores can support inferences about the ability of college students to evaluate information in a digital environment. Eighty-eight undergraduates responded to *iSkills* assessment tasks and to more open-ended "naturalistic" tasks. These naturalistic tasks were designed to be like homework assignments that incorporate the critical evaluation of digital information. We observed weak-to-moderate correlations between scores, suggesting overlap in the skills assessed by the *iSkills* and the naturalistic tasks. Analyses of concurrent cognitive interviews (n=11 of 88) suggested distinctions between students' response processes to the assessment and naturalistic tasks. Although *iSkills* assessment tasks appear to elicit skills consistent with evaluation of digital information in the real world, students' responses to the naturalistic tasks demonstrated broader evaluation skills and less attention to the testing context. This study provides empirical validity evidence regarding ETS's *iSkills* assessment, as well as valuable insights into how undergraduates evaluate information in a digital environment.

INTRODUCTION

There is growing concern among educators and policymakers about the ability of students to critically use digital hardware and software, communication tools, and networks (i.e., information and communication technology - ICT) to meet their information needs. While many of today's college students use a wide array of ICT to achieve a variety of tasks, most seem unable to critically navigate the virtual tidal wave of information caused by the proliferation of ICT throughout academia, the workplace, and society at large (Breivik, 2005, 1998; Oblinger & Oblinger, 2005).

This issue is significant because ICT are ubiquitous in society, particularly in higher education settings, and both the ways in which information is stored, organized, and disseminated, and the literacies needed to access, manipulate, and communicate information are rapidly changing. In today's college classroom it is no longer sufficient to be able to acquire and demonstrate the traditional literacies of reading, writing, and numeracy (i.e., mathematical knowledge and skills). The 21st century college classroom requires students to have strong ICT literacy (ICTL) skills to meet their information needs (International ICT Literacy Panel, 2002; Educational Testing Service [ETS], 2003).

ETS's *iSkills*™ assessment purports to reflect real-world ICT literacy skills. Examinees solve information problems through simulated technology, and these tasks are embedded within scenarios designed to mimic the situations in which college students demonstrate their skill in locating, managing, and using information. Evidence for this validity claim comes from the close collaboration with ICT literacy experts in the design of the assessment, endorsements by additional panels of experts, student exit surveys (Katz, 2007), and empirical comparisons between *iSkills* scores and other assessments (e.g., Katz & Macklin, 2007; Katz et al., 2009). However, objective, empirical validity evidence is limited and additional research still needs to be conducted. The current study uses an argument-based approach to assessment validation to evaluate the extent to which performance on *iSkills* tasks measuring the evaluate performance area can support inferences about the ability of undergraduate students to evaluate information in a digital environment. More details of the study may be found in Snow (2008).

The overarching goal of the work is to investigate whether the *iSkills* assessment tasks and naturalistic ICTL tasks provide comparable measurement of students' evaluation skills. To accomplish this goal, criterion tasks were developed to approximate the context of academic assignments in which undergraduate students are expected to utilize ICT to evaluate information. These criterion tasks were designed to be "naturalistic" representations of how students evaluate information in a technological environment. Student scores and response processes on the *iSkills* evaluate tasks were then compared with their scores and response process on the naturalistic ICTL evaluate tasks.

The naturalistic evaluate tasks differ from the *iSkills* evaluate tasks in two important ways. First, students select their own ICT (within a computer lab setting) to complete the tasks, rather than being limited to using specific web browsers and generic software interfaces (as with the *iSkills* evaluate tasks). Second, the context for the naturalistic evaluate tasks is based on in-depth interviews with undergraduate students about how they evaluate information in a technological environment, as well as actual assignments from college courses in which students have to demonstrate their information evaluation skills.

This study supports the *iSkills* validity agenda by developing ICTL tasks that are closer to real-world equivalents (i.e., are naturalistic) and by examining the extent to which response processes on the *iSkills* evaluate tasks corresponds with response processes on the naturalistic evaluate tasks. Our focus on evidence based on student response processes is consistent with recent research that calls for

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

validity studies that go beyond using correlations as foundational evidence by including a thorough explanation and analysis of how response processes lead, through the attribute(s) being measured, to test scores (Gorin, 2007, 2006; Borsboom, Mellenbergh, & Heerden, 2004; National Research Council, 2001). This study also provides valuable insight regarding how undergraduate students evaluate information in a digital environment.

THEORETICAL GROUNDING

ICT Literacy, Information Literacy, & Evaluating Information

The *iSkills* assessment was designed to assess Information and Communication Technology literacy, the skillful use of information in digital environments. ICT literacy is closely related to concept of information literacy, defined by the American Library Association (ALA) as being "able to recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information" (ALA, 1989), focusing on the information literacy skills as they intersect with use of technology (Katz, 2005). In this work, we focus on a subset of ICT literacy skills associated with the critical evaluation of information.

Much research on information literacy has focused attention on the ways and extent to which information is evaluated (e.g., ALA, 1989; AASL & AECT, 1998; ACRL, 2002), as well as on ways information is evaluated in an ICT-rich environments (e.g., ISTE, 1998; NRC Committee on Information Technology Literacy, 1999; ITEA, 2003, 2000; Partnership for 21st Century Skills, 2003, 2005). This research indicates that a number of factors may be considered in information evaluation, including trustworthiness, relevance, currency, accuracy, objectivity, sufficiency, resource type, and ethical use.

The process of evaluating information has evolved as college students rely on the Internet as their first step in conducting academic research (Friedlander, 2002; OCLC, 2006). In fact, research conducted by the Online Computer Library Center (2006) found that students utilize Internet search engines more than library-specific databases. This broadening of the information landscape necessitates that students be keenly aware of when and how to judge the credibility of information they locate via the Internet (Metzger, 2007). However, do students have this new awareness and can they critically evaluate information in a broader, technology-rich information environment?

Survey results, as well as studies of individual's behavior when conducting research, support skepticism about college students' readiness to critically evaluate information. In an international survey, 70% of college students reported that information is equally trustworthy whether obtained via a search engine or a library website (OCLC, 2006). Research conducted at California State University reports that 28% of surveyed students believe there is a "central Internet authority" that assures the accuracy of Internet information (Manuel, 2002). In a survey of 1,050 college students, almost two-thirds believed that the range of resources on the web was adequate for their needs (OCLC, 2002). Although information search experts might rely on authority and information quality when judging resources (Rieh, 2002), students have been found to judge the usefulness of information based on surface features, such as the density of text on a webpage (Fidel et al., 1999). Students show little understanding of how to differentiate the value of various sources of information (Hepworth, 1999; Caravello, Herschman, & Mitchell, 2001).

The ICT Literacy Framework

In January 2001, the Educational Testing Service (ETS) assembled an international panel to investigate the importance of ICT and its relationship to literacy. The panel's primary tasks were to examine the need for a measure of ICTL in several contexts, including schools, and to develop a framework for ICTL that would provide a foundation for the design of measurement instruments, including large-scale and diagnostic assessments. The panel agreed that little was being done to instruct undergraduate students in critical ICTL skills

(International ICT Literacy Panel, 2002).

Drawing on work by the International ICT Literacy Panel (2002) and the Association of College and Research Librarians (2002), ETS and representatives of seven college and university systems defined ICTL in the higher education context as:

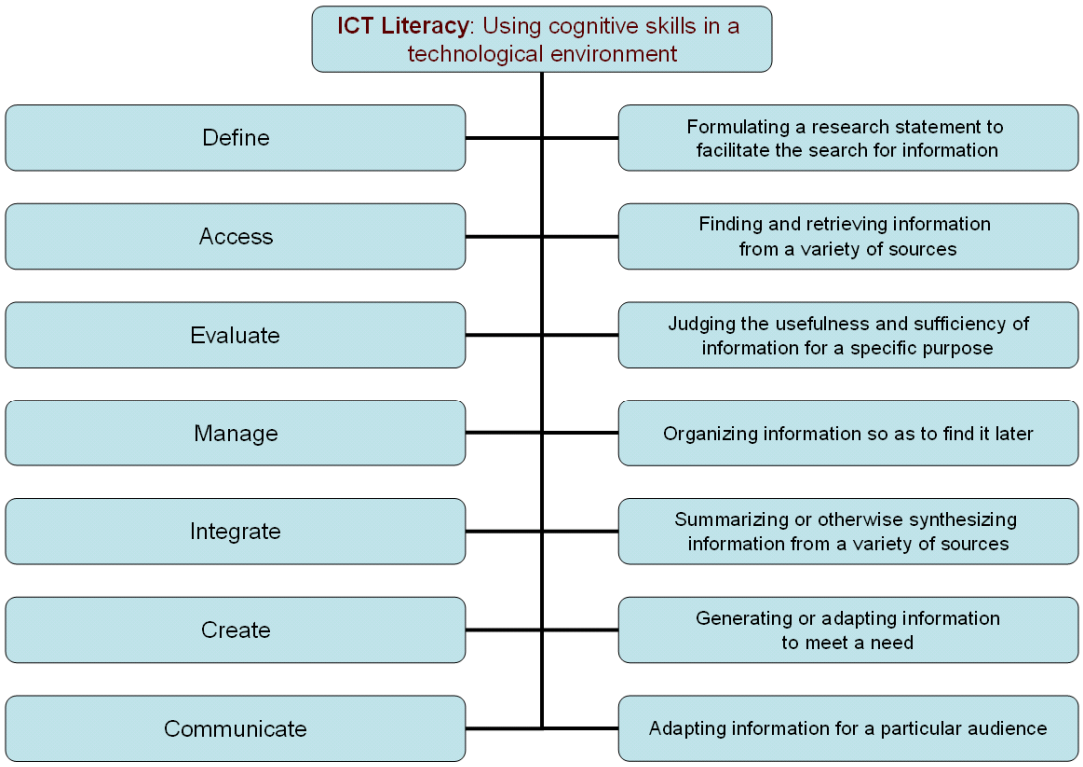
> The ability to use digital technology, communication tools, and /or networks appropriately to solve information problems in order to function in an information society. This includes the ability to use technology as a tool to research, organize, evaluate, and communicate information and the possession of a fundamental understanding of the ethical / legal issues surrounding the access and use of information (ETS, 2003, p. 11).

ICTL is demonstrated by applying cognitive skills in a digital environment to define, access, manage, evaluate, integrate, create, and communicate information. The *iSkills* assessment was designed to measure ICTL through these seven performance areas (Figure 1; Katz, 2007; Katz & Macklin, 2007).

ETS *iSkills* Assessment for Higher Education

The ETS *iSkills* assessment for higher education is administered via the Internet. The assessment consists of scenario-based, information management tasks that simulate real-life situations. Students complete the tasks using a wide array of information and communication technologies, including word processing, spreadsheet, email, file manager, presentation,

FIGURE 1 — HIGHER EDUCATION ICTL FRAMEWORK, EDUCATIONAL TESTING SERVICE



Adapted from "Testing information literacy in digital environments," by I. Katz, 2007.

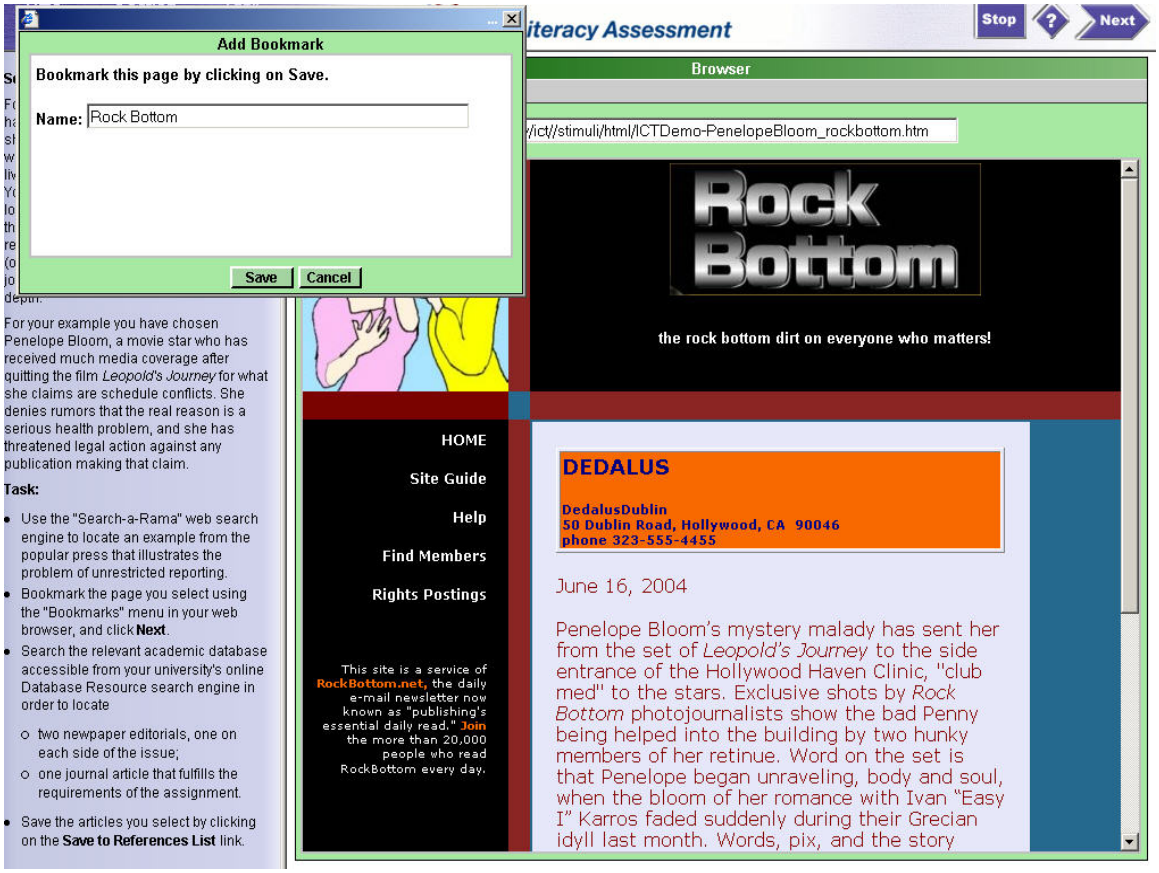
https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

and search engine tools. To avoid bias due to test takers' knowledge of particular software packages, these tools contain generic menu options common to most commercial software packages, but not specific to any (Katz et al., 2004).

The purpose of the *iSkills* assessment is "to determine the degree to which students are sufficiently ICT literate to use digital technology, communication tools, and/or networks to solve information problems likely to be encountered in most common academic and workplace situations" (Katz et al., 2004, p. 9). The assessment was designed to measure student learning outcomes with regards to ICTL. As an "outcomes assessment," a committee of higher education advisors envisioned assessment scores as informing the following: (a) "understanding student ICT literacy, including comparisons of literacy levels between groups of interest," (b) "informing resource allocation at the institution regarding course offerings, such as a basic ICT literacy course, or curriculum content," (c) "advising individual students regarding the potential benefits of enrollment in a basic ICT literacy course," and (d) "advising student preparedness to enter academic years, courses of study, or particular courses based on the level of ICT literacy associated with success in these endeavors" (Katz et al., 2004, p. 9).

Figure 2 shows a screen shot of a sample *iSkills* assessment task that measures students' ability to access and evaluate information. Students

FIGURE 2 — SAMPLE TASK FOR EVALUATING INFORMATION, ETS *iSKILLS* ASSESSMENT FOR HIGHER EDUCATION



From "Testing information literacy in digital environments: ETS's *iSkills* assessment," by Katz, I. R., 2007. Presented at the Alliance for Media Literacy Research Summit, St. Louis, Mo., June 24, 2007.Copyright 2009 by Educational Testing Service.

must search the Internet and an academic database to access different types of resources and then select a subset of resources based on their relative authority, objectivity, and timeliness. Evidence of students' ability to evaluate information comes from the search results that students investigate, as well as the ones they actually select to use in their assignment.

Students' responses are scored on a 3-point scale (0 - incorrect, .5 – partially correct, 1 - correct). The specific raw score is based on the degree to which they visited relevant and trustworthy sites from their search results, accurately determined sufficiency of selected sources, and selected the most appropriate sources (Katz, 2007; Tannenbaum & Katz, 2008).

Interpretive Argument for *iSkills* Evaluate Tasks

The most recent version of the AERA/APA/ NCME Standards (1999) and the latest edition of the book *Educational Measurement* (Brennan, 2006) endorse the view that test validity comprises a process of making an evaluative *argument* that links observed performance with the proposed interpretations and uses of test scores by integrating strands of evidence based on test content, response processes, internal structure, relations to other variables, and consequences of testing.

Kane's argument-based approach to assessment validation (2006, 2004, and 1992; Cronbach, 1988) provides a contemporary and practical model for linking interpretive arguments regarding assessment validity to the evidence needed to evaluate the assumptions and inferences underlying these arguments. Put another way, interpretive arguments specify the reasoning involved in linking observed assessment results to the conclusions and decisions based on the results.

The first assumption in the interpretive argument for the *iSkills* evaluate tasks is that the task content accurately represents undergraduate students' ability to evaluate information in a digital environment (i.e., the ICTL evaluate domain), particularly their ability to judge the usefulness, authority, objectivity (a lack of bias), and timeliness of various types of information sources and, based on these judgments, the extent to which the sources are sufficient for addressing a stated information need. The inferences underlying this assumption that can be evaluated using evidence based on response processes are that (a) undergraduate students respond to tasks with knowledge specific to the ICTL evaluate domain and not other ICTL domains (e.g., integrating information) or extraneous factors such as test taking strategies, and (b) undergraduate students' reasoning for the *iSkills* evaluate tasks reflect the reasoning the tasks were designed to elicit.

The second assumption in the interpretive argument for the *iSkills* evaluate tasks is that the tasks elicit knowledge, skills, and abilities that are consistent with undergraduate students' ability to evaluate information in digital environments (i.e., in the "real-world" outside of testing context). The inference underlying this assumption that can be evaluated using evidence based on relations to other variables and response processes is that performance on the *iSkills* evaluate tasks moderately correspond with performance on the naturalistic evaluate tasks (i.e., the criterion measure of "real-world" performance).

Method

Participants

Eighty-eight undergraduates were administered *iSkills* evaluate tasks and naturalistic evaluate tasks. Of these, 11 students participated in cognitive interviews as they attempted to solve all tasks. Tables 1 and 2 summarize the demographic and academic characteristics of the full group of participants and the cognitive interview group, respectively. The 88 participants were volunteers from a larger group who took a partial form of the *iSkills* assessment (Snow, 2008).

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

TABLE 1—RACIAL/ETHNIC AND GENDER BREAKDOWN, SELECT STUDY ACTIVITIES

	Male	Female	Valid n
Full Group			
White	45	30	75 (85%)
Hispanic	3	1	4 (5%)
Asian American	3	3	6 (7%)
African American	0	0	0
Other	3	0	3 (3%)
Valid n	54 (61%)	34 (39%)	88 (100%)
Cognitive Interviews			
White	7	2	9 (82%)
Hispanic	0	0	0
Asian American	1	1	2 (18%)
African American	0	0	0
Other	0	0	0
Valid n	8 (73%)	3 (27%)	11 (100%)

Note. The "Other" category includes undeclared, Native American, and mixed racial/ethnic students.

TABLE 2—ACADEMIC CLASS AND GENDER BREAKDOWN, SELECT STUDY ACTIVITIES

	Male	Female	Valid n
Full Group			
Freshman	11	8	19 (22%)
Sophomore	15	6	21 (24%)
Junior	8	8	16 (18%)
Senior	20	12	32 (36%)
Valid n	54(61%)	34 (39%)	88 (100%)
Cognitive Interviews			
Freshman	2	2	4 (36%)
Sophomore	2	1	3(27%)
Junior	1	0	1 (10%)
Senior	3	0	3 (27%)
Valid n	8 (73%)	3 (27%)	11 (100%)

Instruments

ETS iSkills Tasks

The partial version of the *iSkills* assessment administered in this study contained eight tasks, each designed to be completed in 3-5 minutes, that measured students' ability to evaluate information and integrate information in a digital environment. This report includes analyses from only the four evaluate tasks. Each task resulted in five scores. Because preliminary analyses suggested redundancy among some scores (Snow, 2008), however, some items were combined and students could earn a maximum of 15 points on the four evaluate tasks.

In two of the four *iSkills* evaluate tasks, students evaluate the quality of several websites resulting from a Google-like search. The "search results" show the URL of the site, its title, and a 1-2 sentence description. For each website, students indicate, by selecting from among provided criteria, whether the site is written by an authoritative source, reflects objective (unbiased) information, and reflects recent information. In the other two tasks, students identify, from among several sources (e.g., websites, journal articles, newspaper editorials), at least two reliable sources that provide opposing viewpoints on a controversial issue (e.g., a public smoking ban). Unlike the previous tasks, students are not explicitly given evaluation criteria and so must decide for themselves how to judge the usefulness (including authority, timeliness, expression of a particular viewpoint) of each source.

Naturalistic Evaluate Tasks

The four naturalistic tasks were designed to measure both information evaluation and information integration skill. This report includes analysis of scores from only the evaluate portions of each task.

Development of the naturalistic tasks proceeded iteratively. Initial design of the tasks followed interviews with 17 undergraduates about how they would evaluate information in the context of hypothetical academic assignments (summaries of actual assignments). The tasks were revised and rubrics developed based on (a) feedback from the first author's dissertation committee and the second author and (b) results of a pilot administration of the tasks (n=18), which included cognitive interviews.

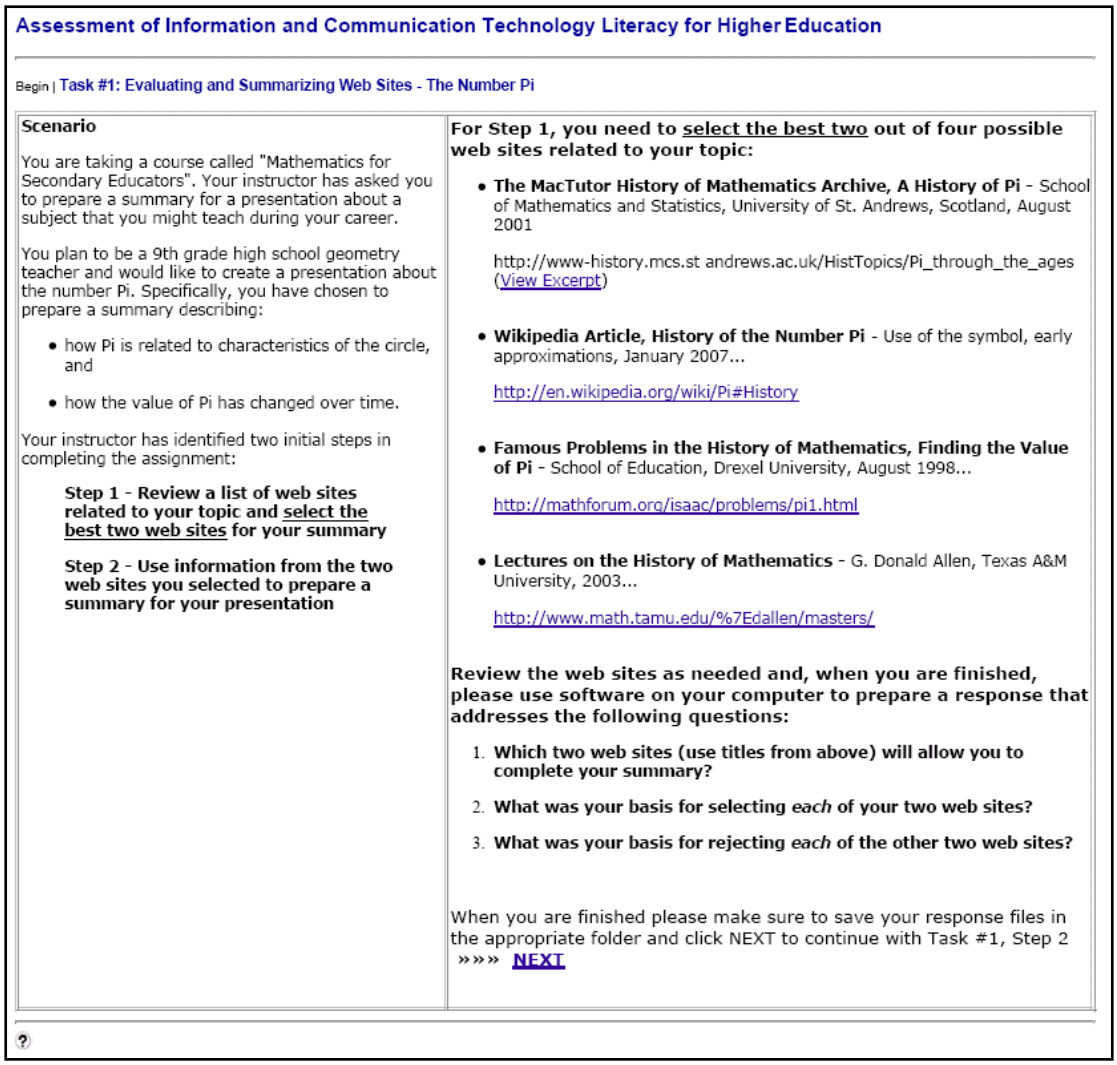
Three naturalistic tasks measured information evaluation skills. Each of these tasks was designed to elicit three scorable observations and take approximately 22 minutes to complete. Each task consisted of an opening academic-based scenario describing one of four possible topics (The Number Pi, Public Smoking Debate, Purchasing Computer, Critical Thinking), as well as several follow-up steps asking students to use ICT tools on their computer to review information sources and describe their basis for selecting and rejecting sources (i.e., evaluate information).

Figure 3 shows a portion of a naturalistic task. The screen is divided into two columns. The left column contains the scenario and steps for completing the task. This column remains stationary as students complete a task so they always have access to the scenario and steps. The right column, however, changes as students complete each step of the task. The right column contains the information sources available for responding to the scenario. The information sources are listed with their title; author/ publisher; publication or retrieval date; and a link to the actual information source, or, if the actual source contains too much information or irrelevant information, an excerpt of the source. Students review the information sources to select the best two sources for completing the task. Finally, students are asked to describe the basis for selecting their two sources and for rejecting the other two sources.

Students have to make several decisions related to evaluating information in order to complete this task: (a) which information sources to review, (b) how to review the sources (i.e., as listed in task or actual source/source excerpt), and (c) which types of information presented with the sources (e.g., author, publication date) form the basis for selecting or rejecting the source. It is the last of these decisions that

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

FIGURE 3—NATURALISTIC ICTL TASK #1 – THE NUMBER PI: STEP ONE, EVALUATING WEBSITES



students are asked to describe; their descriptions are assessed via the scoring rubric.

Procedure

Administration of ETS iSkills Evaluate Tasks

The *iSkills* assessment was administered in a 20-seat campus computer lab. The assessment was delivered on identical PC computers via a secure version of Microsoft Internet Explorer, which ensured that students could not leave the assessment once they had begun. Students received a common set of instructions prior to beginning a background questionnaire. Students were given a total of one hour to complete the background questionnaire and assessment.

Administration of Naturalistic ICTL Evaluate Tasks

The naturalistic tasks were administered in the same 20-seat computer lab as the ETS *iSkills* assessment tasks. The naturalistic tasks were delivered on identical PC computers via student's choice of web browser. Prior to completing the tasks students reviewed a series of web pages containing instructions for completing the tasks and completed a background questionnaire. Students were given a total of two hours to complete the background questionnaire and naturalistic assessment tasks.

Cognitive Interviews

Concurrent and retrospective cognitive

interviews (Willis, 2005; Willis, 1994; Someren, 1994; Ericsson & Simon, 1993) were collected from 11 participants. For the concurrent interviews, a standard set of probing questions (e.g., can you say more?, how did you know that?) encouraged students to talk aloud about how they understood, processed, and responded to the *iSkills* and naturalistic evaluate tasks. For the retrospective interviews, a standard set of probing questions encouraged students to reflect on their solutions and difficulties, and explain why they believed they experienced difficulties with certain tasks (e.g., which tasks were most difficult for you to complete?). Prior to conducting the interviews students were provided with an opportunity to practice thinking aloud. Each interview was recorded and transcribed for subsequent analysis.

Scoring Naturalistic ICTL Evaluate Tasks and Inter-rater Reliability

Figure 4 shows how four factors relevant to the naturalistic ICTL evaluate domain – explicit/ confounded use of expert evaluation criteria, correct/incorrect application of expert criteria, sophistication of expert criteria, and identification of tradeoffs in using expert criteria – are combined into the scoring rubric for the naturalistic evaluate tasks (Snow, 2008, for details on the rubric's development). The left side of the rubric contains the different levels of the four factors arranged into four scored levels. Each level is assigned a numeric value starting with zero for "No Ability to Evaluate Information" and ending with three for "Strong Ability to Evaluate Information." The right side contains example student responses illustrating the four scored levels. Students with no ability to evaluate information (i.e., score of 0) in a technological environment don't use expert criteria, don't correctly apply expert criteria (in any manner), don't use sophisticated expert criteria, and don't identify tradeoffs in using expert criteria when selecting or rejecting information sources.

Middle ability students are able to use more expert criteria, correctly apply the expert criteria, use sophisticated expert criteria, and identify some sophisticated tradeoffs in using expert criteria. High ability students (i.e., score of 3) use two or more expert criteria, correctly apply the criteria, use sophisticated expert criteria, and identify more sophisticated tradeoffs in using expert criteria.

Two raters (first author and an assistant) used the rubric (Figure 4) to score students' responses to the naturalistic tasks. Prior to scoring the entire set of responses, the raters collaboratively scored a random sample of 15 student responses to reach a common understanding of the scoring rubric. The two raters scored the remaining responses independently. Cohen's kappa (κ; Cohen, 1960) for these scores was .84, indicating strong inter-rater agreement.[1] In cases of score discrepancies, the score assigned by one rater (first author) was used.

RESULTS

This section first investigates the statistical relationship between *iSkills* and the naturalistic tasks via inspection of correlations. Next, the main body of this section presents illustrative results from the qualitative analyses of cognitive interviews, comparing and contrasting the approaches taken by a student who scored well on both *iSkills* and naturalistic tasks and a student who scored moderately on *iSkills* but poorly on the naturalistic tasks. By examining and comparing response processes one may be able to infer the degree to which the correlations summarize desired or spurious relationships.

Test Scores

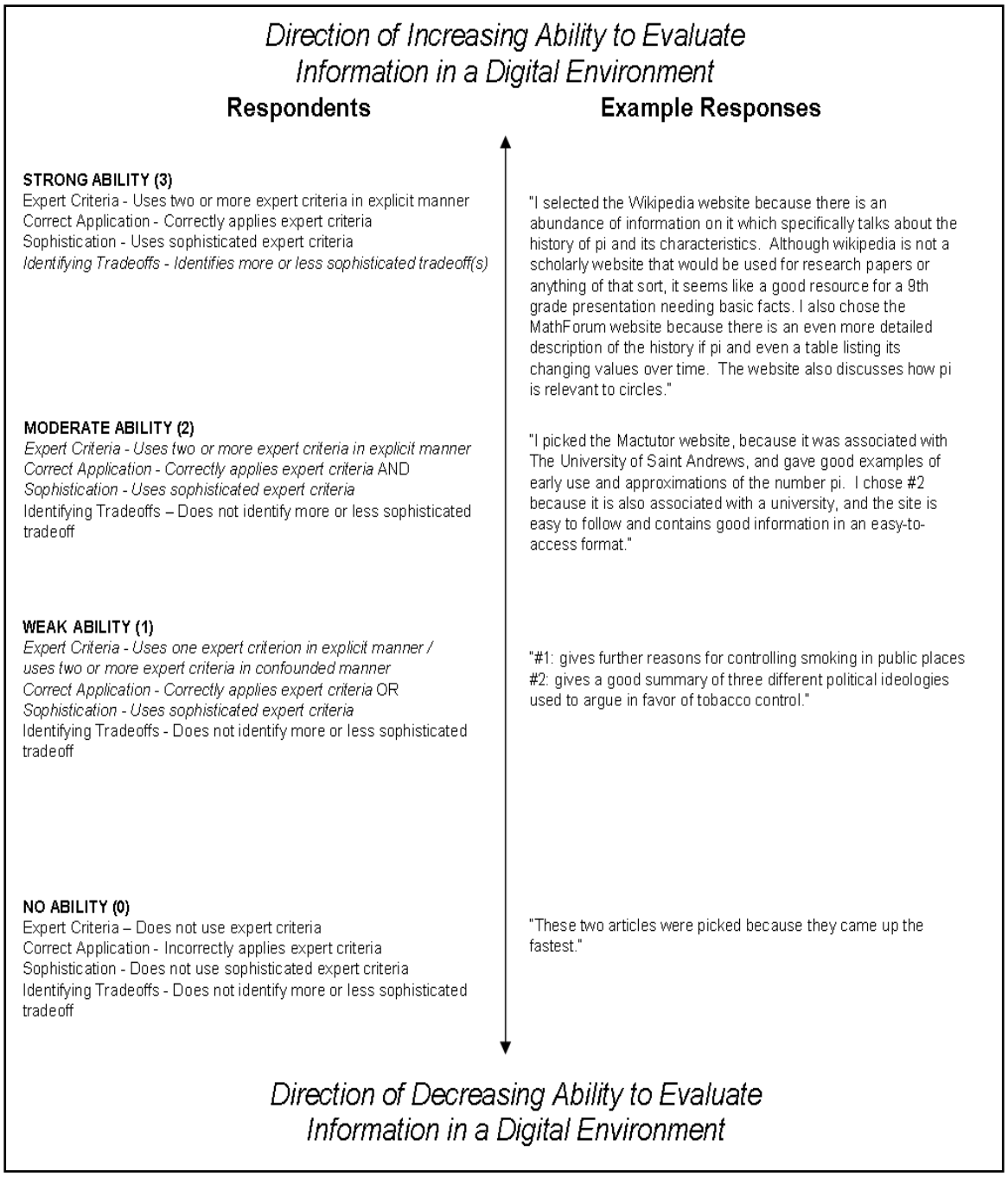
Students performed poorly on the naturalistic tasks, earning on average just 45% of the 15 possible points (M = 6.7, SD = 2.0, n = 88). Naturalistic test scores ranged from 1 to 12. These students did better on the *iSkills* tasks, earning on average 67% of the 15 possible points (M = 10.0, SD = 1.6, n = 88). Scores ranged from 3 to 12. The estimated reliability (coefficient alpha) for the naturalistic tasks was 0.46 and was 0.52 for the *iSkills* tasks.

Correlations

Because of measurement error, the correlations

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

FIGURE 4—SCORING RUBRIC - NATURALISTIC ICTL EVALUATE TASKS



Note. Italics indicate factors that vary from one ability level to the next.

between observed test scores are weaker than the correlations would be if the scores were error free. This reduction is known as *attenuation of correlation* (Spearman, 1904; Lord & Novick, 1968; Zimmerman & Williams, 1997; McDonald, 1999). We use Spearman's correction for attenuation formulas to adjust Pearson correlation coefficients for the impact of measurement error (estimated by the reliability of each instrument) on the *iSkills* assessment scores and naturalistic ICTL assessment scores.

The attenuated (uncorrected) correlation between the *iSkills* evaluate and naturalistic evaluate task scores is 0.19. The disattenuated (corrected) correlation between *iSkills* evaluate and naturalistic evaluate scores is 0.40. The observed disattenuated correlation indicates a weak-to-moderate positive linear relationship between performance on the *iSkills* evaluate tasks and naturalistic evaluate tasks.

Response Processes

This section summarizes the verbal responses to two *iSkills* evaluate tasks and one naturalistic task. Response processes are presented for two students: one who scored above the mean on both *iSkills* and the naturalistic task ("Sally") and one who scored above the mean on the *iSkills* tasks but below the mean on the naturalistic tasks ("Kim"). Kim's case is of particular interest because her performance (and the performance of students like her) may suggest reasons for the weak-to-moderate correlation between scored performance on the *iSkills* evaluate tasks and naturalistic evaluate tasks.

Each student's overall approach is described for three tasks. These tasks were selected because they represent parallel aspects of information evaluation skill:

> 1.*iSkills* T1. In *iSkills* task #1, students evaluate the quality of several websites resulting from a Google-like search. For each site, students must select whether the website meets the evaluation criteria (authority, timeliness, objectivity, relevance) and, if not, on which criterion the website falls short.
>
> 2.*iSkills* T5. In *iSkills* task #5, students evaluate the quality of several information sources without being given explicitly provided criteria. From among these sources (given as citations), students are asked to identify two reputable sources that represent opposing sides of a controversial issue.
>
> 3.NatT1. Naturalistic task #1 is similar to *iSkills* T5 in that students are not provided with criteria to use to evaluate the quality of four information sources and describe their basis for selecting two sources and for rejecting the other two sources (Figure 3).

High Ability Student ("Sally")

Sally received perfect scores across the items embedded in *iSkills* T1 and T5 (Table 3).

iSkills Task 1. Sally's verbal responses to the *iSkills* evaluate tasks reflect her strong ability to evaluate information. She thoroughly reviewed the instructions and anticipated the potential

Table 3—Sally's Scored Responses, iSkills Evaluate Tasks 1 & 5

***iSkills* Evaluate Task**	**Scored Response**				
T1	i1	i2	i3	i4	i5
	Correct	Correct	Correct	Correct	Correct
T5	i1	i2	i3	i4	i5
	Correct	Correct	Correct	Correct	Correct

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

activities, which is more of an expert approach when compared to novices, who tend to cursorily review the instructions and then figure things out along the way (VanLehn, 1989). For example, Sally began *iSkills* Task 1 by reviewing the scenario presented on the first screen:

> *Scenario, evaluating search results… As initial research you've located seven potentially promising websites and need to determine whether these sites are reliable and relevant to your question. Record your evaluation in a table below your web-browser. Select the single most appropriate evaluation of each site based solely on the search results page from the following criteria*

Next, she reviewed the criteria provided in the task and commented on their relevance to the topic presented in the scenario:

> *So…Useful, site is [unintelligible]. Not authoritative, site is not useful primarily because its author is not an authority. Okay, so I'm going to test it for how it just looks, I'm going to look for who created the site, I can look either on the server at the top or maybe it will have something about who wrote it below. Problematically biased, not useful primarily because of the author's attempt to influence the audience. That's pretty given, or the person is just writing with a very biased…which seems kind of weird because I'm talking about [this topic], but maybe. Not current, site not used primarily because information is out-of-date. I think that speaks for itself, something from the seventies doesn't apply much to my current research.*

and anticipated and differentiated the task activities:

> *Okay, after reading this, kind of how I anticipate- determine that these sites are reliable. I feel like what will come up is maybe the first site… what I expect to have happen after this is to tell me something- first of all, this will be over on my left, so I can refer back to it. I think I would expect it to say that 'this is the first site' and give me an example of the browser, then let me search through it, and I will evaluate these things based on that site. Or the other option I can see and I don't know how clear it is here is that it would give me a set of Google search-results where I just get the blurbs with the hyperlinks and that's possible too. I could still judge most of this from those.*

Finally, Sally proceeded to the task screen, attempted to reconcile the actual task with her expectations, and developed a general plan for solving the task:

> *So, I guess I'll let you see what happens. Okay, yeah, so we have a fake web-search, is what I'm guessing, but it's standard and this is what you'd see on any sort of Google, Yahoo, or Dog-Pile sort of search. Okay, so I'm going to look at it and look at the search results it gave me and then try and decide which one is primarily… if it's either useful or I would say it's not and check one of the other three boxes.*

She repeatedly showed a thorough understanding of the value of different types of Internet-based information sources when, in *iSkills* Task 1, she recognized the differences between a blog developed by a teenager:

> *First thing I notice is that it's from a teen-blog. Meaning that somebody is blogging this, a singular person, and they're doing it from personal experience. So from that I can primarily say that this is not a useful site for my research. I would say, primarily, because the author is not an authority. 'Not authoritative'? …. That might not be the best option. 'Not*

> *useful because of the author's intent to influence the audience?' It doesn't seem like there's any sort of obvious bias here. 'Site not useful primarily because the information is out of date?' Well, this is just a week ago, but I think in this case it would be not authoritative. It sounds like the person doesn't have a clear understanding of what it means, in general, to [discuss this topic].*

and a technical publication on a university website:

> *Okay, it sounds more credible, but that doesn't mean that much, because not all universities are credible. The title of the web page sounds more like a thesis. So that automatically makes me think this will have better things than before. Just kind of looking at the website, it's an .edu server, meaning that it has some sort of educational basis, as opposed to the teenblogs.com that I could set up over the weekend if I wanted to. It says 'publication.' It looks like it was fairly recent. It's also in a .pdf format, which, to me, implies more text a larger file size. Probably more good information than you might see on your regular .com page. This is good, it's current, it doesn't seem to have bias, since it sounds kind of like what you're doing, a study of a sample. In terms of whether or not it's authoritative… since it's from a university and nothing about the author, I'm going to assume that, since it was published, according to this, that it's probably good. So I'm going to say useful for this one.*

She identified and applied a wide variety of expert criteria when selecting and rejecting information sources, including relevance, completeness, authority, and timeliness.

iSkills Task 5. Sally began *iSkills* T5 by reviewing the scenario:

> *For a project in class your team needs to find two recently published statements that represent reputable but different viewpoints on a controversial issue. Looking for two recently published statements that have a reputation, but differing viewpoints on a controversial issue… Okay. 'You search a different viewpoints database on the school library website and get the returns listed on the next page.'*

and identifying the context and anticipating and differentiating possible task activities:

> *Okay, so myself, I'm in… a library website, I've got a database, and I'm in this class, that I can forget about, because that information is mostly extraneous. … I'm not so sure what the search result's going to look like for this. 'You need to examine the search returns,' so maybe just the same… we'll see. 'The search engines… in order to determine whether they'd allow your team to complete the necessary research. …reputable but differing viewpoints,' okay.*

Sally proceeded to the next task screen and attempted to reconcile what she saw with her expectations for the task:

> *Okay, just a comment on how it's set up, I've got my search results, and it looks like I will select the one of interest, that I imagine is one of my two viewpoints that I'm interested in and it says what type it is over on the side, an essay or website… it gives me the website. Great, the essays are in journals, so I can look at that too.*

As Sally began to review the first two sources she realized that she was not entirely clear what she as supposed to do with the sources as she evaluated them:

> *It's a magazine, I don't think that's going to have the reputation that's*

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

> *relevant for my class. Oh, okay, so I'm probably going to have to go through and systematically evaluate every one of these. .... It says 'examine them' but I feel like there should be some instructions up here, if this is even what I'm supposed to do, saying 'Select-Select each website and answer the following question' or something more specific, because I guess I don't actually have to select the two right here, it's not asking me to do that, so I could just say 'yes, this one might be good' and click next, then I hope it would save that somehow… then I click this one and I'd say 'no, it wouldn't be good.' I'm just going to give it a shot, because this is the reason I'm doing this.*

She then suggested that an alternative method of presenting the task would make it more clear:

> *So… still, this box should come up at the same time, the second box should, and it should have an arrow, not necessarily for the yes part, but 'select the materials above that appear to be most helpful in meeting the assignment.' Because when you see this thing your inclination is to start clicking the boxes. I didn't even know I was clicking the boxes in the beginning, I thought I was clicking the boxes to evaluate this question. This part needs to be more pertinent. The only reason I'm clicking the boxes is to say 'If it does work, here are the ones that I would use.' That was kind of ambiguous. I feel like this box needs to come up at the same time. 'Does the database allow your team to complete the necessary research?' Without the boxes. Yeah, that would work better. No boxes, the primary question. It's also nice because it doesn't make it so we have to do so much multitasking mentally.*

Sally appeared to have formed expectations about the correct method for responding to the tasks. When she came to a different type of evaluate task she initially struggled to reconcile her prior method with the method(s) required to complete this new task:

> *Because you see the boxes, and also, this as a prerequisite to other tasks we've done, we've been trained through tasks one through four to click boxes and answer questions about what we've clicked. This is different, it needs to have a different interface. I'm just going to pretend they're not even here to proceed through the rest of this. - I guess my point is that I was able to get there eventually, it was just not clear, it could be clearer. I don't think they're trying to test that, I think they're trying to test how well I understand the good and bad points of all these different types of sources. I would hate to think that somebody would do poorer on this test merely because the instructions were a little bit ambiguous at times.*

Once Sally worked out a procedure for responding to the task she returned to reviewing and comparing each of the sources listed in the reference list against the requirements of the task. For example, she rejects one source because, based on the source (a magazine), title, and website URL, it did not seem reputable:

> *I don't think the second one is one I'm interested in, because it's from a magazine. [reads magazine tagline] Clearly biased. Using what we learned from those last web interfaces, that's just biased and I'm going to discount that as something not useful. Recently published statements, reputable but differing viewpoints – not reputable. Plus it's a website, I mean, [name].com- no, no, no, no.*

In another example, Sally selects a source because, based on the source (a journal), title,

and date of publication, it is relevant, somewhat reputable and timely:

> *Okay, perfect. Here's the other side. It's a journal, recent, , sixteen pages. She's not a doctor or anything, but it has an abstract, and it says it's an opinion article, it's an opinion issue, there's no legislation about it, so it seems like its okay.*

Note that Sally is initially somewhat cautious about the source being an opinion article from a possibly non-authoritative author, but is willing to select the source because the task calls for differing viewpoints and there is no legislation clarifying what is legal and illegal vis-à-vis the research topic stated in the scenario.

In the end, Sally correctly judged that the available sources were sufficient and selected two sources as representing reputable, but different viewpoints on the topic.

Naturalistic task 1. Sally received scores of 2 (out of 3 possible points) on both of her written responses to the two evaluate items embedded in NatT1 (Table 4).

Sally began NatT1 by reviewing the scenario and anticipating and differentiating possible task activities and tools (i.e., software programs):

> *Task 1. So it tells me what task I'm on like I said it would. Evaluating and summarizing websites… subheading on the number 'pi'. Okay. So probably going to be something about- It will give me some website links maybe and I'll have to go through and just summarize them like it says about the number pi. Which will be fun, because pi is interesting. I'll read my scenario first 'You're taking a course called 'mathematics for secondary educators'' Okay, so for secondary education…. 'You plan to be a ninth grade high school geometry teacher and would like to create a presentation about the number pi. Specifically you have chosen to prepare a summary.' Okay. Ninth-grade teacher, presentation… right off the bat I think PowerPoint because it's a pretty effective way of presenting things. … Okay, so I have two goals for my project, and my instructor has two*

TABLE 4—SALLY'S WRITTEN RESPONSES, ITEMS 1 & 2, NATURALISTIC ICTL TASK 1

Naturalistic ICTL Task	Score	Written Response
NatT1i1 - Source Selection	2	The first website was from an educational server (from the .edu) and was actually published information. It concisely presented the information that accomplished my second goal of understanding how Pi changed throughout history. The second website came from Wikipedia, and is presumably trustworthy content since it is on a web page that will probably get a lot of hits from internet users. In addition, it offers diagrams and information on both the history of Pi and its relationship with circles.
NatT1i2 - Source Rejection	2	The third site was hosted on a math forum website, which means that the creater could have no experience or validity when they made the website. It is a .org site, yet it offered little specific information I wanted for my presentation. The fourth and final website was probably full of credible information, yet the majority seemed to not pertain to my topics or aims for this particular presentation. It was more difficult to navigate and not specific to the topic of interest.

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

> *steps for it. 'Step 1:' I need to 'review a list of websites related to your topic and select the best two.' Okay, so I'm going to get a list of hyper-links and they'll probably say website 1,2,3 whatever, for organization I guess. Then I'll select the best two… find some way to keep track of those, maybe in the write-up. I don't know, we'll see. 'Use information from the two websites you selected to prepare a summary for your presentation.' Okay, 'Please click 'next' to work on Task 1 Step 1.' So I'm going to get a list of websites here, and hopefully I'll be able to click around and navigate through them and kind of see which one… I'll have my scenario on the left still…good, then I'll be able to check back and forth.*

Sally proceeded to the next screen, reviewed the instructions for Step One and began to develop a strategy for evaluating the information sources and responding to the questions:

> *…So I've got… my four websites and 'a review of the websites is needed and when you are finished please use software on your computer to prepare a response that addresses the following questions.' Okay- so now I'm starting to kind of multi-task in my head here. I guess it's… I want to kind of go through what I have to do and then I'll look through to the websites and find the two that best do my goals. That's where I'm going to stop there. I'm not going to do anything with them except maybe copy and paste their links somewhere, into a random Word or Notepad or something so that I can keep track of them. Then… then it sounds like before I even do the presentation I'm going to open up a Word document and answer these questions here for you. Then it'll probably go on to step two to have me make the presentation. That being said….*

With a basic strategy in-hand Sally began to review the available information sources in more detail. She skimmed the information provided for each source and determined the overall value of the source for preparing a presentation about the number Pi. Once Sally determined each source's overall value she compared each source against one another and against the information needs of the task. In the following example, Sally reviewed two of the available information sources, *Wikipedia Article, History of the Number Pi* and *The MacTutor History of Mathematics Archive, A History of Pi*:

> *As untrusting as Wikipedia can be… generally, on topics that are more general, like pi or like the civil war they've been reviewed by a lot of people- so if something was wrong with it, it would have probably been reported and fixed. It kind of depends on what it all looks like….I don't know who the author is, but I don't know who the author is on Wikipedia either… kind of a good sign of telling how much a site is used on Wikipedia is all of the different links and places it travels in terms of the network. … I think I would choose Wikipedia in this case, just because it seems to have a little bit more information and equation work. Generally I have a no-use rule for myself on Wikipedia, but that's for things like research papers… this is just math calculations… I don't know…. So for my ninth graders, and because it's not something, like, persuasive. I'm not trying to- I'm not implementing laws here, I'm not trying to be creative and come up with a solution- that's what I'd use peer reviewed sources for. I'm just trying to get facts…. Wikipedia, as I said, is Wikipedia, but it offers a wealth of information right now, and since it's on something like pi it's been looked at by a lot of people and is probably pretty well edited to this point. Plus, the stuff I'm taking is pretty basic."*

Once she skimmed the sources, she compared them against the goals of Step One of the task:

> *So both- what was my second thing I needed to be looking at? How the value of pi has changed over time and how it's related to a circle? This had a good history…Where's… how it relates to a circle. Honestly, how- Yeah, okay. This is kind of where I would be looking at that. The circle area, pi r squared… the diagram does a nice job, I would probably copy and paste that into a presentation because it's pretty colors and kids might be able to understand kind of what's happening, even at a ninth-grade level. Ah! Here's the geometry! A lot of how it relates to circles, spheres, ellipses… cylinders which are built out of two circular bases going up… cones which has a circular base going up…so that has quite a bit of good information… just because I know a little bit about, I can tell you right now that this is correct, pi r squared, two pi r, four thirds pi r cubed… yeah. Yeah, same thing for a cylinder…So… it looks like a pretty good website.*

and against one another:

> *In terms of- now it's like I kind of want to compare these two websites. Because they both seem alright in terms of facts. This one came from Wikipedia…history…St.Andrews… I know it's from the UK… it sounds kind of like a school… but it could also just be, like, a project on a server. I can't really tell… the school of mathematics and statistics, Scotland. Okay, so probably pretty credible. I don't know who the author is, but I don't know who the author is on Wikipedia either… beyond… um…it has a lot of… kind of a good sign of telling how much a site is used on Wikipedia is all of the different links and places it travels in terms of the network. Everything is connected via a whole bunch of- like a spider-web model… so if it has a whole lot of extra links like this at the bottom it means that people have gone to this and gone other places and looked around.*

In the end, Sally had a difficult time choosing between *Wikipedia Article, History of the Number Pi* and *The MacTutor History of Mathematics Archive, A History of Pi*. She prefered the information in the *Wikipedia* source over the *MacTutor* source, but struggled to select the former when creating her summary in Step Two of the task. Importantly, she recognized that *Wikipedia* tends to be inappropriate for research papers, but that the context for the task – preparing a presentation on the number Pi for 9th grade students – may allow it to be used as a primary source:

> *Again- it seems pretty good…ew…I don't know which one I'd decide- for me personally, maybe just because I know a little bit more about it- although this is simpler and smaller, I think I would choose Wikipedia in this case, just because it seems to have a little bit more information and equation work. Generally I have a no-use rule for myself on Wikipedia, but that's for things like research papers… this is just math calculations… I don't know.*

Later, once she reviewed all of the available sources, Sally continued to reflect on the context of the task as she went through the process of considering different evaluation criteria and making final selections:

> *So that being said, this had both, and this one was really good for history. The history of it, not so much how it related to the circle at all. So for my ninth graders, and because it's not something, like, persuasive. I'm not trying to- I'm not implementing laws here, I'm not trying to be creative and come up with a solution- that's what I'd use peer reviewed sources for. I'm*

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

> *just trying to get facts. So I'd use these two, these first two [MacTutor and the Wikipedia].*

When asked to explain her selections in more detail, Sally wavered between *The MacTutor History of Mathematics Archive, A History of Pi* and *Lectures on the History of Mathematics*, primarily because she saw the sources as containing similar information and as generally credible:

> *That one… MacTutor? It's just hard, because it's not like before where it's, like, the credibility of a peer-reviewed source versus just an online site. But again- you don't really need that for this one. So I guess I won't let myself get so hung up on it. Informationally, these two seem like probably the best, for me. Yeah- my conflict is between the first one and the third one. Like I said, I can pretty much discount the fourth one. What kind of gets me is because I'm not sure if there's a right answer. It's a forum… it looks like it's written by- and it's associated with the school, it's not like someone just made one up, I could write a forum in a few hours… it has organization and a table of contents. So it sounds like this is probably credible. Wikipedia, as I said, is Wikipedia, but it offers a wealth of information right now, and since it's on something like pi it's been looked at by a lot of people and is probably pretty well edited to this point. Plus, the stuff I'm taking is pretty basic. This one… did a nice job with swinging the history.*

In order to make a final selection Sally cross-referenced the competing sources against *Wikipedia Article, History of the Number Pi*, which was the source she already selected. Sally ended up selecting *The MacTutor History of Mathematics Archive, A History of Pi* because the source excerpt contained a full citation,[2] which, to her, indicated that it had been published and had an extra degree of credibility over the alternative source:

> *But I didn't actually cross-reference these, but I could kind of compare this to Wikipedia I suppose. I'd hope I'd kind of look at it… I mean, it talked a little bit about the Bible and pi equals three… it said that in both Wikipedia and this one…it talked about the Babylonians in this one… My conflict is just between…I know Wikipedia is one that I want to use, my conflict is between these other two, and it's just because both of them seem like possibly credible sources…Here's a source…yeah. Yeah, yeah, yeah… published. Okay- oh! That helps, this is published, and this is not. That makes my choice easier. Because they're kind of on the brink, but that allows me to do the first two. MacTutor and Wikipedia.*

Low Ability Student ("Kim")

Kim got all five items correct in iSkills T1 but only two out five items correct in iSkills T5. She did not get any items in iSkills T1 and T5 partially correct (Table 5).

Kim's verbal responses to the *iSkills* evaluate tasks suggest her lower evaluation ability. Unlike Sally, Kim only superficially reviewed

TABLE 5—KIM'S SCORED RESPONSES, ISKILLS EVALUATE TASKS 1 & 5

***iSkills* Evaluate Task**	**Scored Response**				
T1	i1	i2	i3	i4	i5
	Correct	Correct	Correct	Correct	Correct
T5	i1	i2	i3	i4	i5
	Correct	Incorrect	Incorrect	Incorrect	Correct

the scenarios, forming limited expectations for the activities she may have to engage in as she responded to the tasks.

iSkills Task 1. Instead of analyzing the scenario as Sally did, Kim focused immediately on the information sources, stated criteria, and questions:

> *Yeah, I just started at the top, read through here, and then went down to the tasks, and focused more on what qualified rather than just the top paragraph. Just because of the context. In the test- When you're test-taking, usually the scenario doesn't matter a whole lot. It's more the criteria and how you're supposed to rank things, it's more those. ...I don't imagine them asking me too many questions about [the topic]. More about ranking each of the search results.*

Kim proceeded to the next task screen and conducted an initial review of the reference list:

> *So now…I'm looking over here, where the browser and the search results popped up. I'm thinking I'm probably going to read through them all. First I read the titles, the headings, just because they're in blue, and they pop out a little bit more. I'll probably read through all the headings… the [first heading] because it qualifies… I'll look over here at the qualifications. That one's probably a useful site because it has the university's name on it. Then… [the second heading] sounds pretty good also.*

As Kim reviewed the reference list she noticed and then began to focus on the sources and criteria as they were presented in the table below the reference list: "I'm more of a visual learner, so tables are great, and they summarize a lot of information. I really like tables." At this point, Kim, like Sally, began to iteratively compare the information sources in the reference list in more detail against the criteria available in the table. For example, Kim rejected the second source because she determined that it wasn't timely. She selected another source as useful because it appeared to be on topic and was likely produced, or at least published, by an authoritative organization:

> *The description is pretty good, there's mention of a national symposium … the web address is also a .gov address, which suggests it's a little more authoritative… This is a toughy. I would probably mark this one as useful.*

Kim completed this task by describing how she made her final selections in the table below the reference list:

> *I went down, I like to go down the list and look at all the options before I make a choice, because I'm indecisive, but I just like to look at and pick out the best sounding one. Try to pick that one out first, and then maybe a second best in terms of usefulness, out of the list. Then going back, the ones that were blatantly, like the advertising one I went to third. Made it problematically biased, and then whatever was left over that I hadn't selected in the list, I went back and looked at those and just reviewed the useful and made sure that I thought that was useful and went back. Double checked kind of. I thought this one would be useful too.*

Note that Kim reviewed and selected sources primarily through a process of elimination. She initially focused on identifying and selecting the best sources and then, based on these selections, moved to identifying why the remaining sources were not useful.

iSkills Task 5. Kim began *iSkills* T5, as she did with *iSkills* T1, by paraphrasing the task objectives:

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

> *So I read through the description of what you need to do. So you need to sort out websites and figure out which ones… which ones are useful and which ones are not, but they don't give you the criteria list like they did in number one and three. So it sounds like the same thing, but they're not telling you what applies as useful or not.*

Note that Kim explicitly compared this task against the requirements of previously completed *iSkills* evaluate tasks. Unlike Sally, Kim next reviewed the entire screen before reviewing each source in the reference list, "I'm just looking at the whole screen, then I look through this little box down here because that's what they're asking. So it's asking 'Does it allow you to complete the necessary research?' Then I'll go through and read the headings." As Kim began to review the sources in the reference list she reviewed the scenario again and realized that she, like Sally, was not entirely clear why there were check boxes next to each information source in the reference list:

> *Some of the titles are pretty funny. So what it's asking for? I'll go and review what it's asking for, you're supposed to write. I don't know what these little boxes are for…I don't know if I'm supposed to check them or not. I'll just leave them unchecked, and answer the question.*

Kim began reviewing the available sources by looking at the headings and taking note of where each source originated from (e.g., journal or website):

> *I'll look at the headings, and at the same time look at where they are coming from. Four of them are from journals, and two of them are websites. It tells me. This one seems more useful than not because they're from journals…I'm looking at them, actually, more closely. Two are from journals, and two are from, just magazines. So… and two are just websites. So websites… to me, websites probably aren't that- some websites should be taken that seriously, just because anybody can have a website. So… these two, I would probably label as not useful, then…let's see…*

At an early stage in the task Kim separated websites and journals as containing different levels of authority and possibly accuracy. Specifically, Kim suggested that, because anyone can have a website, they may be less authoritative than journals. Based on this perspective, Kim focused her attention on reviewing the titles of the four journal sources provided in the reference list to determine the extent to which they were relevant to the specific task requirements:

> *I'm just double-checking that they're all kind of on the topic that I'm supposed to be writing on… so- they all sound like they're pretty much on topic. Actually, I'm mainly looking at these four right here. Right here. The first two and then the… three and four, or four and five, sorry. I'm just looking at the titles and seeing if they apply to what I'm supposed to write about and it sounds like they all do apply, or would be useful. So I'm going to put, yeah, they do.*

Note that, at this stage, Kim only reviewed the titles of each source and did not or could not provide specific details about why, exactly, she believed that the four journal sources were relevant and sufficient for completing the necessary research.

Once Kim decided that the available sources were sufficient for completing the task she began to review all of them, including the previously "rejected" websites, in more detail. For example, she selected two of the available sources, but for different reasons:

> *So this one would be interesting to see- it's from a major newspaper which is a pretty reputable source, …, it would be interesting to read though just because it did get printed. Let's see, this one… Let's see this one is talking about a related topic. I'm just reviewing the topic. Yeah it would probably be on topic because it deals with a related issue.*

Specifically, she selected the first source because it came from a major newspaper, which she saw as a reputable source. On the other hand, she selected the other source because it talked about an issue she sees as being relevant to the topic of the task.

In another example, Kim selected a third source because she saw it as being possibly relevant to the task topic. Kim focused on the degree to which the source title included mention of the main topic and mentioned that she would need to look at the source in more detail to determine its relevance to the task scenario. Importantly, Kim completed the task by selecting three sources, rather than the two required by the task.

Naturalistic Task 1. Kim received scores of 1 and 2 on her written responses to items 1 and 2 in NatT1, respectively (Table 6).

Kim began NatT1 by paraphrasing the task requirements:

> *So it's asking to look at a list of websites and choose two, um, that allow it to how pi is related to characteristics to a circle and how it's changed over time and then in step two you want to make a summary for your presentation from the two websites that you choose.*

She then reviewed the available sources by looking at their titles. She also reviewed the entire right side of the screen to identify the specific activities she needed to complete:

> *Okay…so reading the titles…And looking down underneath the titles, just to see what it's asking me to do… more specifically. It's asking for which ones and the basis for selecting them and the basis for rejecting them so I go back up to the websites…And try to do number one which is try to find the best two. Um…so I'm reading the titles again. And descriptions.*

Next, Kim reviewed the four available sources in more detail. For example, she rejected the first source, *Wikipedia Article, History of the Number Pi*, immediately because it can be changed by anyone with access to the Internet. For her, this made the source unreliable and possibly non-authoritative:

> *Uh, probably wouldn't use a Wikipedia article. Just because it's… you know what Wikipedia is, and*

TABLE 6—KIM'S WRITTEN RESPONSES, ITEMS 1 & 2, NATURALISTIC ICTL TASK 1

Naturalistic ICTL Task	Score	Written Response
NatT1i1 - Source Selection	1	I selected these two websites through the process of elimination. The two sites I chose also appear to directly address the topics I need to summarize.
NatT1i2 - Source Rejection	2	I rejected the Wikipedia site because of its ability to be edited by anyone. The sources were not cited on the site. I rejected the History of Math website because there were many links to different topics, none of which sounded like they were directly addressing Pi.

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

> *anybody can go on there and change it. Um…Wikipedia a share site so you can…there's different little topics and you can, anybody can go on and change the topic but, um, usually if something's wrong it gets corrected pretty quickly. But if there's a common misconception about something and it gets put on there it usually stays on there a little bit. So there was also a pretty funny cartoon of like…somebody was typing about HeMan on Wikipedia and there was Skeletor on the other end…*

Note that Kim qualifies her rejection somewhat by stating that the reliability and authority of *Wikipedia* as a source depended on the type of information being described. She believed that topics that are well-researched and known have a "self-correcting" nature in *Wikipedia* in that errors are quickly corrected. At the same time, common misconceptions about certain types of information can be proliferated. Importantly, even though Kim recognized the potential of *Wikipedia* as a source for certain topics, she was not willing, at least initially, to select it as one of her two sources to use in the second part of the task.

Kim identified the remaining three sources as generally relevant to the task topic because their titles all reference history and mathematics in some way:

> *Um…let's see…the other ones, the other three that are left…there is 'The History of Math', 'Problems in Math' and now I'm going back over to column one as a reminder that I need to find the characteristics of the circle and how it's changed. So the history one is also pretty good. It looks like all three of them…the three that are left are about history.*

Kim began to review each of the remaining three sources in more detail by clicking on the available links. Specifically, Kim focused on whether or not Pi was mentioned as one of her primary criteria for selecting or rejecting a source. For example, using this as criteria to review the source, *Lectures on the History of Mathematics*, allowed her to skim it for relevance rather than reading it in detail:

> *So…looking at all the titles because you probably don't need to read all of them because I'm just looking for Pi. But it doesn't…I'm just kind of skimming it and I don't see them mentioning Pi anywhere so…I'm going to go back. And I'm probably not going to choose this one because I don't have time to go through it all.*

Once Kim finished reviewing the sources, she briefly returned to the task requirements before choosing Microsoft Word to describe her selections and rejections, as well as the rationale behind each:

> *So, I wanna pick…we're finished… um, so do I…? Let's see. Oh, I have to read the instructions. To use software on this computer. Yeah I'll just use Word because I have a PC at home. Which two websites use the titles…so I picked. I'm just going to copy and paste…the 'Mac Tutor History of Math.'*

When asked, Kim described her rationale for selecting *The MacTutor History of Mathematics Archive, A History of Pi* as being a process of elimination. Specifically, she already comfortably rejected *Wikipedia Article, History of the Number Pi* and *Lectures on the History of Mathematics* sources, which left her with only two sources to choose from:

> *I chose that one…I think I chose that one…maybe I didn't choose that one. No I did. Um, that one I chose because I didn't choose the other two.*

Similarly, Kim selected *Famous Problems in the History of Mathematics, Finding the Value of Pi* because it was easy to access the information (i.e., contained tables), was relevant

to the task topic, and, through a process of elimination, was the only source left in the reference list:

> *Oh, this one I chose because…it's a pretty good site, all of the info seems fairly relevant. I like the table. And I didn't choose the other ones. Elimination, again because I didn't want the Wikipedia one or the lecture on history because it seemed too broad.*

DISCUSSION

Recall that Kane's argument-based approach to assessment validation provides a model for linking interpretive arguments regarding assessment validity to the evidence needed to evaluate the assumptions and inferences underlying these arguments. This section examines the extent to which evidence presented in this report (correlation and response processes) supports the interpretive argument for the *iSkills* evaluate tasks.

The first assumption of the interpretive argument is that the *iSkills* evaluate tasks elicit responses consistent with the definition of the ICTL evaluate domain (the intended construct) and not of other ICTL domains or irrelevant factors. The verbal responses for Sally and Kim demonstrated two cases in which students who performed well on *iSkills* tasks responded to those tasks using knowledge specific to the ICTL evaluate domain and not other ICTL domains, and that the tasks elicited the reasoning they were designed to elicit (evaluation skill). For *iSkills* Task 1, Sally and Kim applied given evaluation criteria, including usefulness (relevance), objectivity (bias), authority, and timeliness, when selecting and rejecting information sources. Their verbal responses also indicated that they understood the value of different types of Internet-based information sources—they recognized the differences between a blog developed by a teenager and a website ending in .org or .edu, as well as the differences between the quality of information contained in a website, newspaper, and journal article. For *iSkills* T5, Sally and Kim identified and applied similar expert criteria to evaluate the available information sources, and correctly determined that the sources were sufficient for fulfilling the stated research need. Note that the verbal responses showed a difference between Sally and Kim in their propensity to respond to the *iSkills* evaluate tasks using test-taking strategies. For *iSkills* T1, Kim, who scored below the mean scale score on the naturalistic evaluate tasks, identified the given criteria as more important to completing the task than the scenario. She also applied the given expert criteria to the available information sources through a process of elimination, rather than selecting and rejecting the information sources based on a detailed analysis of each source. These test-taking strategies are similar to those that students use to respond to forced-response (i.e., multiple-choice) items. Thus, *iSkills* task #1 may lead some students to adopt strategies not consistent with evaluation skill as they appear in real-world tasks in which selection-by-elimination is not feasible. This observation suggests that future revisions of *iSkills* evaluate tasks focus on developing more non-criteria-given tasks. Not only might this decrease the likelihood of students with good test-taking strategies scoring higher on the *iSkills* evaluate tasks, it would also increase the tasks' authenticity (i.e., in real-world tasks, students are not often provided with the criteria they need to evaluate information sources).

The second assumption of the interpretive argument is that *iSkills* tasks elicit knowledge and skills consistent with real-world, digital information evaluation skill. Both the correlation with naturalistic tasks and the verbal responses provide some support for the argument. The observed correlation of 0.19 between the *iSkills* evaluate scores and naturalistic evaluate scores (the criterion measure of "real-world" evaluation skills) indicates a weak association between the two types of tasks. However, given the relatively high degree of measurement error present in both measures, this result does not necessarily indicate that the *iSkills* evaluate tasks are a weak measure of "real-world" ICTL evaluate skills;

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

when measurement error is taken into account the correlations between the disattenuated scores increases to 0.40, which indicates a weak-to-moderate association between performance on the *iSkills* evaluate tasks and naturalistic evaluate tasks.

The evidence based on response processes indicates that the observed correlation, while not ideal, represents desired, rather than spurious, relationships. The verbal responses indicated that undergraduate students responded to the *iSkills* evaluate tasks with knowledge and reasoning both specific and irrelevant to the ICTL evaluate domain. Responses from the *iSkills* evaluate tasks moderately corresponded with verbal responses from the naturalistic evaluate tasks. In both *iSkills* and naturalistic tasks, students demonstrated knowledge of information sources (e.g., title, date), information types (e.g., journal articles, web pages), and expert criteria (e.g., relevance, authority, timeliness) considered. The verbal responses to both measures indicated a difference in whether students considered the extent to which the information sources were complete vis-à-vis that stated research need (i.e., contained the information needed to fully respond to the research need).

CONCLUSIONS

The evidence discussed above moderately supports the interpretive argument for the *iSkills* evaluate tasks. Assessment validation, however, is an ongoing process and the evidence presented in this paper is intended to add to, but not establish, the body of validity evidence for the *iSkills* assessment. Each inference drawn from test scores needs to be evaluated using one or more types of validity evidence. The meaning of constructs, particularly new constructs such as the ICTL evaluate construct, can shift, which results in new inferences that need to be evaluated, often using new types of validity evidence. With regard to the *iSkills* assessment, there are at least three areas in which further validation research could be conducted.

First, the sample on which this study's findings were based was relatively small. Replicating the study with a larger and heterogeneous sample of undergraduate students may help strengthen some of the findings and conclusions, as well as increase the likelihood of their generalizability. In particular, we expect that a larger sample would strengthen the correlation between scores on the *iSkills* evaluate tasks and the criterion measure (due to smaller measurement error) and would help clarify the response processes underlying the observed correlations.

Second, this study provides an example of how to conduct appropriate validation research for a performance-based measure such as *iSkills*. In particular, the study demonstrates how to develop a criterion measure (i.e., naturalistic evaluate tasks) for a new construct (i.e., ICTL evaluate skill). While the evidence indicated that the naturalistic evaluate tasks were a reasonable measure of undergraduate students' ICTL evaluate skills, there were several aspects of the tasks that could be improved before they could be considered a *criterion* measure of "real-world" ICTL evaluate skills. For example, evidence based on response processes indicated that the naturalistic tasks elicited skills related to, but outside of, the ICTL evaluate domain. Students were given access to the entire information source in the naturalistic evaluate tasks to increase authenticity; having access to the entire information source, however, appeared to cause students to consider the sufficiency and completeness of the information source, both skills that could be considered as falling outside of the ICTL evaluate domain. As another example, one of the more challenging aspects of developing a scoring rubric (Figure 4) for the ICTL evaluate construct was deciding how to best arrange the factors into a hierarchy, starting with the most complex understanding of evaluating information in a digital environment and ending with the least complex understanding. It is possible that problems with the rubric (e.g., rubric score categories were not comprehensive or appropriately ordered), rather than problems with the naturalistic evaluate tasks, resulted in the low test ceiling (i.e., few high scores), low internal consistency of the

naturalistic evaluate task scores, and, ultimately, the weak-to-moderate correlation between scores on each test.

Third, there is a need to further specify students' knowledge and skills associated with the ICTL evaluate construct, as well as for other ICTL constructs (e.g., integrating information). Ethnographic methods (interviews, participant, non-participant observation; see Spradley, 1980; Wolcott, 1995; LeCompte & Schensul, 1999) could be applied in academic and non-academic settings to provide a rich and up-to-date understanding of the knowledge and skills associated with ICTL constructs. Classroom observations could be conducted to examine the presence and variation in information literacy instruction and performance in secondary and post-secondary settings. Such work would be of particular importance given the novel and relatively undefined nature of the ICTL constructs. Further research in this area might lead to more valid item designs and scoring rubrics for ICTL constructs, create a stronger foundation of evidence for developing criterion measures of ICTL constructs, and clarify the knowledge and skills used by students to deal with information in 21st century digital environments.

NOTES

1. Cohen's kappa is a commonly used measure of rater agreement that accounts for agreement due to chance and ranges from 0 (no agreement) to 1 (complete agreement). Unweighted Kappa was used.

2. Excerpts of information contain full citations of their original source. Information that is presented in its original context (i.e., not as an excerpt) does not contain a full citation, as students are expected to obtain this from reviewing the actual source.

REFERENCES

American Association of School Librarians (AASL), Association for Educational Communications and Technology (AECT). (1998). Information literacy standards for student learning. In, *Information power: Building partnerships for learning*. Retrieved December 2003 from http://www.ala.org/

American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, D.C.: AERA.

American Library Association (ALA) Presidential Committee on Information Literacy. (1989). *Final report*. Chicago, IL: ALA.

Association for College & Research Libraries. (2002). *Information literacy competency standards for higher education*. Retrieved December 2009, from http://www.ala.org/ala/mgrps/divs/acrl/standards/standards.pdf.

Borsboom, D., Mellenbergh, G., & Heerden, J. V. (2004). The concept of validity. *Psychological review*, 111, pp. 1061–1071.

Breivik, P. S. (2005). 21st Century learning and information literacy. *Change Magazine*.

Breivik, P. S., & Senn, J. A. (1998). *Information literacy: Educating children for the 21st century* (2nd ed.). Washington, DC: National Education Association.

Brennan, R. L. (Ed.). (2006). *Educational Measurement* (4th ed.). Westport, CT: American Council on Education, Praeger Publishers.

Caravello, P. S., Herschman, J., & Mitchell, E. (2001). Assessing the information literacy of undergraduates: Reports from the UCLA library's information competencies survey project, ACRL Tenth National Conference. Denver, CO.

Cronbach, L. J. (1988). Five perspectives on the validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3–15). Hillsdale, NJ: Lawrence Erlbaum.

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

Educational Testing Service (ETS). (2003). *Succeeding in the 21st Century: What higher education must do to address the gap in information and communication technology proficiencies*. Retrieved December 2009 from http://www.ets.org/Media/Tests/Information_and_Communication_Technology_Literacy/ICTwhitepaperfinal.pdf.

Ericsson, K. A, & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Fidel, R., Davies, R. K., Douglass, M. H., Jenny, K. et al. (1999). A visit to the information mall: Web searching behavior of high school students. *Journal of the American Society for Information Science and Technology, 50*(1), 24–37.

Friedlander, A. (2002). *Dimensions and use of scholarly information environment. Introduction to a data set assembled by the Digital Library Federation and Outsell, Inc.* Washington, DC: Digital Library Federation and Council on Library and Information Resources. Retrieved December 20, 2007 from http://www.clir.org/pubs/reports/pub110/contents.html.

Hepworth, M. (1999, August). A study of tertiary students' information literacy and skills: The inclusion of information literacy and skills in the undergraduate curriculum. Paper presented at the 65th IFLA Council and General Conference, Bangkok, Thailand.

Gorin, J. (2007). Reconsidering issues in validity theory. *Educational Researcher, 8*(36), 456–462. Washington, DC: American Educational Research Association.

Gorin, J. (2006). Test design with cognition in mind. *Educational measurement: Issues and practice, 4*(25), 21–35.

International Technology Education Association (ITEA). (2003). *Advancing excellence in technological literacy: Student assessment, professional development, and program standards*. Available from ITEA website, http://www.iteaconnect.org/Publications/publications.htm.

International Technology Education Association (ITEA). (2000). Standards for technological literacy: Content for the study of technology. Available from ITEA website, http://www.iteaconnect.org/Publications/publications.htm.

International ICT Literacy Panel (2002). *Digital transformation: A framework for ICT literacy*. Princeton, NJ: Educational Testing Service. Retrieved November 12, 2007, from www.ets.org/ictliteracy/digital1.html.

International Society for Technology in Education (ISTE). (1998). *National Educational Technology Standards for Students (NETS·S)*. Retrieved June 2003 from http://cnets.iste.org/students/s_stands.html.

Kane, M. (2006). Validation. In R.L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). American Council on Education, Praeger Publishers.

Kane, M. (2004). Certification testing as an illustration of argument-based validation. *Measurement, 2*(3), 135–170.

Kane, M. (1992). An argument-based approach to validity. *Psychological bulletin*, (3), 527–535.

Katz, I. R. (2007). Testing information literacy in digital environments: ETS's *iSkills*™ assessment. *Information Technology and Libraries, 26*(3), 3–12.

Katz, I. R. (2005). Beyond technical competence: Literacy in information and communication technology. *Educational Technology Magazine*, *45*(6), 44–47.

Katz, I. R., Elliot, N., Attali, Y., Scharf, D., Powers, D., Huey, H., et al. (2009). Multiple methods of assessing information literacy: A

case study. *ETS Research Spotlight,* Issue 2, 21–27.

Katz, I. R., Williamson, D., Nadelman, H., Kirsch, I., Almond, R., Cooper, P., et al. (2004). *Assessing information and communications technology literacy for higher education*. Paper presented at the 30th Annual Conference of the International Association for Educational Assessment (IAEA), Philadelphia, PA, June 13–18, 2004.

Katz, I. R. & Macklin, A. S. (2007). Information and communication technology literacy: Integration and assessment in higher education. *Journal of Systemics, Cybernetics and Informatics, 5*(4), 50–55.

LeCompte, M. D., & Schensul, J. (1999). *Designing and conducting ethnographic research (Book 1, Ethnographer's Toolkit)*. Walnut Creek, CA: AltaMira Press.

Lord, F. M., & Novick, M. R, (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Manuel, K. (2002). Teaching information literacy to Generation Y. *Journal of Library Administration, 36*(1/2), 195–217.

McDonald, R. (1999). *Test theory: A unified treatment*. New Jersey: Lawrence Erlbaum Associates.

Metzger, M. J. (2007). Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology, 58*(13), 2078–2091.

National Research Council Committee on the Foundations of Assessment; Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.) (2001). *Knowing what students know: The science and design of educational assessment*. Washington, D.C.: National Academy Press. Available at: http://www.nap.edu/books/0309072727/html/

National Research Council Committee on Information Technology Literacy. (1999). *Being fluent with information technology*. Washington, D.C.: National Academy Press. Retrieved January 2004 from http://www.nap.edu/html/beingfluent/

Oblinger, D. G., & Oblinger, J. (2005). *Educating the net generation*. Boulder, CO: Educause.

Online Computer Library Center [OCLC] (2006, June). College students' perceptions of libraries and information resources. Retrieved December 20, 2007, from http://www.oclc.org/reports/pdfs/studentperceptions.pdf.

Online Computer Library Center [OCLC] (2002, June). OCLC white paper on the information habits of college students. Retrieved February 2, 2008, from http://www5.oclc.org/downloads/community/informationhabits.pdf.

Partnership for 21st Century Skills. (2005). *Assessment of 21st century skills: The current landscape*. Tucson, AZ: Author. Retrieved September 2005 from http://www.21stcenturyskills.org/images/stories/otherdocs/Assessment_Landscape.pdf

Partnership for 21st Century Skills. (2003). *Learning for the 21st Century: A report and mile guide for 21st century skills*. Tucson, AZ: Author. Retrieved December 2009 from http://www.21stcenturyskills.org/downloads/P21_Report.pdf.

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology, 53*, 145–161.

Snow, E. (2008). *Can college students evaluate information sources? Validating a web-based assessment of evaluation skills*. (Doctoral dissertation, University of Colorado, Boulder, 9 August 2008). Ann Arbor: UMI, 2008. 3315820.

https://pdxscholar.library.pdx.edu/comminfolit/vol3/iss2/6
DOI: 10.15760/comminfolit.2010.3.2.75

Someren, M. W. (1994). *The think aloud method: A practical guide to modeling cognitive processes*. San Diego: Academic Press.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72–101.

Spradley, J. (1980). *Participant observation*. Orlando: Harcourt Brace College Publishers.

Tannenbaum, R. J. & Katz, I. R. (2008). *Setting standards on the core and advanced iSkills assessments* (ETS RM-08-04). Princeton, NJ: Educational Testing Service.

VanLehn, K. (1989). Problem-solving and cognitive skill acquisition. In M. Posner (Ed.), *The foundations of cognitive science* (pp. 527–580). Cambridge, MA: MIT Press.

Willis, G. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks: Sage.

Willis, G. (1994). *Cognitive interviewing: A "how to" guide*. North Carolina: Research Triangle Institute.

Wolcott, H. F. (1995). *The art of fieldwork*. Walnut Creek, CA: AltaMira Press.

Zimmerman, D. W. & Williams, R. H. (1997). Properties of the Spearman correction for attenuation formula for normal and realistic non-normal distributions. *Applied Psychological Measurement, 21*, 253–270.