Evaluating A Representative State Sample of Connecticut Seventh-grade Students'
Ability to Critically Evaluate Online Information

Elena Forzani and Cheryl Maykel
University of Connecticut

Abstract

This study investigated the extent to which a sample of seventh grade students (n = 591) in Connecticut critically evaluated online information both within and across three different assessment formats. The formats included Closed (simulated Internet environment requiring constructed responses), Open (actual, unrestricted Internet environment requiring constructed responses), and Multiple Choice. Results indicated that critical evaluation was more difficult for students than the three other online reading and research skill areas assessed (i.e., Locate, Synthesize, and Communicate) in all three formats combined, and was one of the most difficult of the skill areas within each of the three formats. Additionally, among the four critical evaluation tasks assessed (e.g., finding out the author of a website, determining if that author is an expert, evaluating the author's point of view, and evaluating the overall reliability of a website), evaluating the author's expertise and evaluating the overall reliability of a website was the most difficult for students. Finally, students performed better on critical evaluation tasks in the Multiple Choice format than they did in either of the two performance-based formats. Findings suggest that critical evaluation persists as one of the most difficult online comprehension and research skills for students, especially when measured in a performance-based format.

Evaluating A Representative State Sample of Seventh-grade Students'
Ability to Critically Evaluate Online Information

The new Common Core State Standards (2012) that Connecticut has adopted call for students to "assess the credibility and accuracy" of a variety of digital information sources (p. 41). This means that today's students must become proficient not just at gathering information sources and using them to produce writing, but also at evaluating them first to determine their relevancy and accuracy for the task at hand. As more and more of the texts students read and use move online, this skill becomes increasingly important for readers.

Digital information sources, like the Internet, have necessitated the use of new literacy skills as well as new ways of thinking about traditional literacy skills (Coiro, Knobel, Lankshear, & Leu, 2008; Lankshear & Knobel, 2006), such as source evaluation. Students today must learn how to conduct online research and comprehend various types of online texts if they are to be successful both with the Common Core standards and in today's digital world (Common Core State Standards Initiative, 2012; Organisation for Economic Co-operation and Development & the Centre for Educational Research and Innovation, 2010). When using online information, higher-level skills, such as critical evaluation (CE), become especially important (Goldman, Braasch, Wiley, Graesser, & Brodowinska, 2012), since anyone can publish to the Internet (Cope & Kalantzis, 2000; Fabos, 2008). Therefore, the reader, rather than a publisher, bookseller, or other intermediary, becomes the first, and, in many cases, only judge of the accuracy and reliability of information.

Many may assume that today's students are skilled at effectively collecting and communicating reliable online information, since they have, presumably, used the Internet for much of their lives. However, this assumption may not be accurate. Although adolescent

"digital natives" (Prensky, 2001) may be skilled with texting, gaming, social networking, creating mash-ups from multiple media sources, and downloading video and MP3 files, they are not always as skilled with the use of online information, and especially with the CE of online information (Bennett, Maton, & Kervin, 2008; Sutherland-Smith, 2002; Wallace, Kupperman, Krajcik, & Soloway, 2000). In fact, adolescents often overgeneralize their ability to read and research online information effectively because they are skilled with other online and tech-related tasks (Grimes & Boening, 2001; Kuiper, 2007).

Success in conducting research online often is dependent on the reader's evaluation of the information found (Goldman et al., 2012; Wiley et al., 2009). As students search for and synthesize information from various sources, CE skills help guide their decisions about the accuracy of that information (Goldman, Braasch, Wiley, Graesser, & Brodowinska, 2012). A recent study showed how valuable CE skills are to an online research task. Goldman and colleagues (2012) found that students who were more skilled in CE were more focused and efficient (Goldman et al., 2012). When students searched for and incorporated information from various sources, CE skills guided their decisions about the accuracy of that information, helped them to determine what information to use from each source, and informed them of what to look for next (Goldman et al., 2012). As the Internet becomes increasingly central to full participation in today's society, the critical evaluation of information found online becomes more important for both students and educators to understand.

Perspectives and Theoretical Background

The current study is framed by both a dual level theory of New Literacies (Coiro, Knobel, Lanskshear, & Leu, 2008; Leu, O'Byrne, Zawilinski, McVerry, & Everett-Cacopardo, 2009) and by perspectives on the critical evaluation of online information, especially the reliability of

sources. It builds on previous work to investigate how well students in Connecticut critically evaluated online information both within and across three different assessment formats.

New Literacies: A Dual Level Theory

As the pace of technology change accelerates, so too does the pace with which literacies change. The literacies we use in our everyday and working lives are thus always continuously new. This poses a challenge for educators, who must keep up with the many new literacies available to their students. Some (Leu, O'Byrne, Zawilinski, McVerry, & Everett-Cacopardo, 2009; Leu, Kinzer, Coiro, Castek, & Henry, 2013) have thus proposed a dual level theory of New Literacies to address the challenge of conceptualizing literacies that are constantly changing. This theory conceives of literacy as having two interacting levels: an uppercase *New Literacies* and a lowercase *new literacies*. Uppercase New Literacies are broader, more stable, and consist of multiple, integrated perspectives. Lowercase new literacies are more rapidly changing and are comprised of more specific tools, such as text messaging (e.g., Lewis & Fabos, 2005), or of focused disciplinary areas, such as the semiotics of multimodality in online media (e.g., Kress, 2003). The frequent changes occurring within new literacies are guided by the broader, uppercase New Literacies, just as New Literacies are expanded upon and informed by changes within the specific contexts of the lower case literacies.

A commonality across uppercase New Literacies is that the Internet facilitates the advent of new online social practices (lowercase new literacies) that use lower case technologies, such as instant messaging, wikis, blogs, email, search engines, and social networks (Greenhow, Robelia, & Hughes, 2009; Lankshear & Knobel, 2006). The assessment used in this study was situated within a social network environment that required students to interact with student avatars through instant messages, emails, and wikis in the process of completing a research task.

The assessment was thus informed by the uppercase concept that acknowledges the importance of online social practices while at the same time utilizing many lower case new literacies and acknowledging that online social practices occur with the use of many different tools.

The new literacies of online research and comprehension (Coiro, 2003; Leu, et al., 2011) is one of many lowercase theories. This theory seeks to describe what happens when we conduct research and read online. It suggests that at least five processing practices occur during online research and comprehension with a complex layering of both traditional and new skills and strategies that appear in several areas: 1) reading to define important questions or problems (Leu, Kinzer, Coiro, & Cammack, 2004); 2) reading to locate information (Bilal, 2000; Guinee, Eagleton, & Hall, 2003); 3) reading to evaluate information (Sanchez, Wiley, & Goldman, 2006); 4) reading to synthesize information (Goldman, Wiley, & Graeser, 2005; Leu et al., 2013; Jenkins, 2006); and 5) reading and writing to communicate information (Greenhow, Robelia, & Hughes, 2009). Within these five areas reside the skills, strategies, and dispositions that are both important for offline reading comprehension and also distinctive to online research and comprehension. This creates an interaction of both old and new literacies that we are still seeking to fully understand.

In the current study, we used both levels of New Literacies theory to frame our investigation. An uppercase theory of New Literacies suggests that new, online social practices have become important. Online research and comprehension, one of several lower case theories of new literacies, suggests that locating, evaluating, synthesizing, and communicating information are important areas to consider when we conduct research and read online. Thus, we evaluated students' ability to locate, evaluate, synthesize and communicate information within an online research task that required students to engage in several social practices using text

messaging, wikis, email, search engines, and a social network. We focused particular attention in this study on the evaluation of online information, specifically the evaluation of author, point of view, and reliability of source.

Critical Evaluation

The critical evaluation of online information is one of the most important skill sets required by readers today (Goldman, et al., 2012; Wiley et al., 2009). Yet, it is often the area of online research and comprehension with which students struggle the most (Kuiper & Volman, 2008). Lower-level skills, such as locating information on the Internet, may be easier for students to master than higher-level skills, such as evaluating the source and reliability of information. Thus, students may acquire and use information without having the skills to effectively evaluate its accuracy (Grimes & Boening, 2001). Moreover, students may overestimate their ability to critically evaluate online sources (Grimes & Boening, 2001). Students who are less skilled at determining the quality of information and who merely locate information without strategically evaluating it may end up falling behind their more savvy peers, who have the skills to effectively evaluate information before deciding whether and how to use it.

Research on critical evaluation has focused on a variety of information quality markers (e.g., accuracy, authority, comprehensiveness, coverage, currency, objectivity, reliability, and validity), but it often condenses these markers to credibility and relevance as the two main constructs (Judd, Farrow, & Tims, 2006; Kiili, Laurinen & Marttunen, 2008). This study focused on the credibility of the author or source of a website, defined in terms of expertise (Bråten, Strømsø, & Britt, 2009; Judd, Farrow, & Tims, 2006; Rieh & Belkin, 1998), and on the evaluation of the reliability of information (Goldman, et al., 2012; Kiili, Laurinen, & Marttunen, 2008; Sanchez, Wiley, & Goldman, 2006).

Much of the previous research on critical evaluation has focused on college students' abilities (Bråten, Strømsø, & Britt, 2009; Goldman, et al., 2012; Sanchez, Wiley, & Goldman, 2006). This research has had an important impact, leading to critical evaluation and higher-level thinking becoming important components of the recent Common Core State Standards (2012) in the U.S. This research also has had a similar impact on frameworks for K-12 education in other nations such as the recent Australian Curriculum (Australian Curriculum Assessment and Reporting Authority, n.d.). While our understanding of college-aged students' ability to evaluate information, especially online information, has gained greatly from this work, we know much less about younger students' ability to critically evaluate online sources. Given that this is now part of many nations' curriculum frameworks, it is an important area of inquiry. Teachers need to know students' current capabilities as they begin to plan for and teach these important aspects of curriculum.

Thus, this study sought to determine how well students in Connecticut performed on a measure of critical evaluation compared to three other online research and comprehension skills: locating, synthesizing, and communicating online information. This study also evaluated how well students performed in four different aspects of critical evaluation. Two of these were related to the credibility of the author or source of a website, defined in terms of expertise (Judd, Farrow, & Tims, 2006; Rieh & Belkin, 1998), and two were related to the evaluation of the reliability of information (Bråten, Strømsø, & Britt, 2009; Goldman, et al., 2012; Kiili, Laurinen, & Marttunen, 2008; Sanchez, Wiley, & Goldman, 2006).

Specifically, this study evaluated seventh grade students on their ability to: 1) identify the author of a webpage; 2) evaluate the author's expertise; 3) identify the author's point of view; and 4) evaluate the overall reliability of the webpage. This study also sought to determine how

well students performed on critical evaluation in three separate assessment formats, including a closed Internet assessment context (Closed, a simulated Internet environment), an open Internet assessment context (Open, the actual, unrestricted Internet), and a multiple choice context. While all three formats followed similar research scenarios, only the Closed and Open formats were performance-based, and most directly represented an actual online research experience.

Method

Participants

This study is part of a larger study that sampled seventh-grade students in two states in the northeastern United States. The present study, however, reports on the results of a representative sample of students' performance from only one of these two states. A total of 19 school districts were included in the sample, with one participating school per district. In each school, one teacher with two classes of approximately 20 students participated. In a few smaller schools, it was necessary to include two teachers with one class of approximately 20 students each. Districts and schools were selected using stratified random sampling. The sampling plan stratified schools according to three factors: 1) district percentage of Free and Reduced Price Lunches, (a proxy measure of socioeconomic status); 2) performance on the state reading comprehension assessment; and 3) geographical location (rural, urban, and suburban). This was done while taking note of school size. Schools were randomly sampled within each of these strata.

Principals at each of the selected schools identified the English Language Arts teacher or teachers (in the case of smaller schools) whose students best represented the school population and who were willing to participate. Teachers then selected two of their classes that best fit this description. Students from the selected classrooms who had parental consent and who gave their

assent were allowed to participate in the ORCA assessments. This included a total of 725

seventh graders. Each student was assigned to complete one assessment activity on each of two

days. The majority of students completed both of the planned assessment activities. However,

due to absences and a few system errors, 18.5 percent of the sample did not complete both

activities. Thus, the final sample for the present study included 591 students.

Online Research and Comprehension Assessments (ORCAs)

Eight research scenarios were developed using eight different life science topics, all

requiring students to read and conduct research online. Each of these scenarios was developed in

three different formats that included Closed, Open, and Multiple Choice (see Table 1). The

Closed format allowed students to conduct their research in a closed online environment. This

environment was created so that students could search for, select, and use websites from the

project's search engine, "Gloogle," which was only populated with a predetermined set of

websites. The Open format allowed students to search for, select, and use websites from the

actual, Open Internet using Google. The Closed and Open formats were thus largely

performance-based measures. Finally, the Multiple Choice format confined students to selecting

sites and answers from a set of four answer choices per question. Each question and answer set

was accompanied by screenshots of the websites or other web tools (e.g. emails, wikis) that

students needed to use in order to successfully answer the questions. Students could toggle

between the different screenshots as needed by clicking on various links or tabs. The Multiple

Choice format thus attempted to provide students with a richer context than traditional multiple

choice assessments.

In all scenarios, students were presented with science research problems that focused on

the domain of health and human body systems, an area common to many seventh grade science

curricula, with each of the eight scenarios focusing on a different topic. All topics are listed in Table 1. The scenarios were framed around two types of research: "Learn More About (LMA)" and "Investigate Conflicting Claims (ICC)." Half of the scenarios presented the research problem to students via an email message from the school board president (LMA scenarios) and half via a class wiki with a message from the teacher (ICC scenarios). LMA scenarios asked students to learn more about the research topic and to form a main idea about what they learned. ICC scenarios, on the other hand, asked students to investigate two sides of an issue and to take a position (See Table 1).

Each scenario included items assessing students' ability to locate, evaluate, and synthesize information during their research. The scenarios also included items assessing students' ability to communicate the results of the research via either email or wiki. Each scenario, called a LESC, represented each of the four skills areas of Locate, Evaluate, Synthesize, and Communicate with 16 score points per LESC and 4 score points per skill area. Each score point evaluated an online research and comprehension skill identified both from previous research and from discussions with researchers in this area. Each skill area (Locate, Evaluate, Synthesize, and Communicate) included three process skills and one product skill, with one score point assessing each skill, for a total of four score points in each of the four LESC skill areas.

The LESC questions appeared within a Facebook-like environment through avatars named Brianna and Jordan, who were introduced as students from another school. The questions did not appear in a linear sequence according to skill area. Rather, a more natural and logical sequence was used according to the nature of the research task. Students were guided to engage

in the four skill areas through their online research tasks via requests and questions from Brianna and Jordan.

Table 1

The Eight LESC Scenarios by Topic

Topic	Research Question	Type of Research	Communication Tool Used in the Research
Energy Drinks	How do energy drinks affect heart health?	Learn more about	Email
Heart-Healthy Snacks	How do snacks affect heart health?	Learn more about	Email
Volume Level	Can listening to volume levels on an MP3 player cause hearing loss?	Learn more about	Email
Ringtones	How well can adults hear mosquito ringtones?	Learn more about	Email
Third-hand Smoke	Is third-hand smoke dangerous to lung health?	Investigate conflicting claims	Wiki
Asthma	Can Chihuahua dogs cure asthma?	Investigate conflicting claims	Wiki
Contact Lenses	Do cosmetic contact lenses harm your eyes?	Investigate conflicting claims	Wiki
Video Games	Do video games harm your eyes?	Investigate conflicting claims	Wiki

Note. Each Topic was developed in three different formats that included Closed, Open, and Multiple Choice.

The four score points for CE related directly to three of the traditional critical evaluation criteria that include authority, objectivity, and accuracy (Judd, Farrow, & Tims, 2006; Rieh & Belkin, 1998; Bråten et al., 2009; Goldman, et al., 2012; Kiili, Laurinen, & Marttunen, 2008;

Sanchez, Wiley, & Goldman, 2006). Students were prompted by Jordan to determine the author of a given website (authority), evaluate the author's expertise (authority), identify the author's point of view with a supporting detail (objectivity), and evaluate the overall reliability of the site using at least one piece of valid reasoning (accuracy). The responses for the four CE score points were obtained through an instant message conversation with the avatar Jordan, who prompted students to access a website at a provided link. From the website, students had the opportunity to navigate to the author biography page, which was hyperlinked to the given site. If students navigated to the biography page, they then had the opportunity to gather more information on the author to inform their responses. However, students were not directly asked to navigate to the biography page, and the link appeared somewhat differently in different LESCs, depending on the site that was used. Therefore, not all students accessed the additional information, and responses varied greatly.

Scoring the ORCA

An auto-capture system recorded students' responses for all score points for later scoring. Video screen captures recorded students' performance as a backup for the auto-capture system, and to score search activities that occurred outside of the assessment system in the Open Internet format. Three process score points and one product score point were calculated for each of the four major skill areas (Locate, Evaluate, Synthesize, and Communicate) using a binary (1 or 0) score point system. Each student completed two LESCs, so each student's final score was comprised of an overall total of 32 score points.

The Multiple Choice reports were scored automatically by the ORCA scoring system. However, the Closed and Open reports were hand-scored by a team of eight scorers, with one scorer assigned to one of the eight topics each. Scorers were trained by two expert scorers to a

minimum inter-rater reliability level of 90% accuracy for each score point. Each scorer was then released to score his or her LESC topic. Throughout the scoring process, the scoring of each score point was checked using a random sample of 20 student reports by one of two expert scorers within each set of 100 reports scored (20% of all Closed and Open assessments). Scorers who did not continue to meet 90% accuracy for each score point, within each set, were retrained and retested to this level before continuing scoring.

Procedures

LESC Administration.

The ORCAs were administered during two assessment days held at each school. Before testing began, students were assigned to assessment topics and formats following a specific assignment plan that was designed to ensure equal and random assignment of students from various schools across LESCs. They were then entered into the ORCA database and assigned a unique identification number by the system. On each assessment day, students were read brief, standardized instructions before beginning the ORCAs, which used an automated start-up sequence on a set of MacBook Airs. By entering their unique ORCA identification numbers into the login screen, students were brought directly to their assigned ORCA in the online system. Students who typically received accommodations in the classroom received the same accommodations during the ORCA assessments. The test administrators for the ORCA were two graduate students from the university who, together with the lead Investigator, developed a protocol for school set up and test administration.

Scoring Procedures.

The operational definition for each score point was similar across all three formats of the ORCA: Closed, Open, and Multiple Choice. However, the scoring process differed slightly for

each format. For the Closed and Open formats, score reports were generated by the data capture tool of the ORCA system for each completed LESC and were used to score the Closed and Open formats, with one exception. In the Open condition on Synthesis tasks, QuickTime videos were used to score the Locate questions since the auto capture system could not capture students' searches on the open Internet.

Analysis.

Analysis of variance (ANOVA) was used to answer all four of the present study's research questions: 1) How well do seventh-grade students, in all three formats combined (Closed, Open, and Multiple Choice), perform on critical evaluation compared to three other online research and comprehension skills (locating, synthesizing, and communicating)?; 2) How well do seventh-grade students perform in four dimensions of critical evaluation, including identifying the author of a webpage, evaluating the author's expertise, identifying the author's point of view, and evaluating the overall reliability of the webpage?; 3) How well do seventh-grade students perform in each format separately on critical evaluation compared to three other online research and comprehension skills, including locating, synthesizing, and communicating information?; and, 4) How well do seventh-grade students perform on critical evaluation, comparatively, in each of the three formats?

Results

Prior to the statistical analysis, all data were examined and found to meet assumptions of analysis of variance (ANOVA), including repeated measures ANOVA. A bonferroni correction was used to control for Type I error when conducting all post-hoc comparisons. To investigate the first research question, a one-way repeated measures ANOVA was conducted to compare students' scores in each of the four skill areas in all three formats combined. Multivariate

statistics revealed that there was a significant effect for LESC skill area score, Wilks' Lambda = .418, F (3, 588) = 272.76 p <.0005, multivariate partial eta squared = .582. An analysis of pairwise comparisons showed that there was a significant difference between each of the four skill areas and each other skill area (p < .05 for all pairwise comparisons). Students' scores were highest in Synthesize (M = 6.07, SD = 1.81), followed by Locate (M = 4.52, SD = 2.21.), Communicate (M = 4.22, SD = 2.28) and, finally, by Evaluate (M = 3.61, SD = 1.88). Thus, students scored the lowest on Evaluate.

Table 2

Student Performance by LESC Skill Area Within and Between Three Formats

	Locate M (SD)	Evaluate M (SD)	Synthe-size M (SD)	Comm-unicate M (SD)	Statistical Test M (SD)	Effect Size M (SD)
All Three Formats**	4.52 (2.21)	3.61 (1.88)	6.07 (1.81)	4.22 (2.28)	F (3, 588) = 272.76	n_p^2 = .58
Closed only	3.85 (2.27)	2.84 (1.54)	6.32 (1.76)	3.12 (1.86)	F (3, 191) = 327.44	multivariate n_p^2 = .84
Open only	4.44 (2.32)	2.71 (1.43)	6.06 (1.86)	3.00 (1.74)	F (3, 167) = 209.14	multivariate n_p^2 = .79
Multiple Choice only	5.15 (1.87)	4.95 (1.67)	5.87 (1.78)	6.07 (1.67)	F (3, 224) = 43.19	multivariate n_p^2 = .37

Note. p < .05

To address the second research question, a second one-way repeated measures ANOVA was conducted to compare students' scores on the four Evaluate skills. The means and standard deviations are presented in Table 3. Multivariate statistics showed a significant overall effect for all four Critical Evaluation score points, Wilks' Lambda = .390, F (3, 588) = 306.950, p < .0005, multivariate partial eta squared = .61. An examination of pairwise comparisons showed that there was a significant difference in student performance between each of the four score points

and each other score point (p < .0005 for each pairwise comparison), except between score point 2 (evaluating author expertise) and score point 4 (determining the overall reliability of a website). Score point 1 (determining the author of a website) had the highest mean (M = 1.62, SD =.61), followed by score point 3 (determining the author's point of view and providing supporting evidence; M = .77, SD = .77), score point 2 (determining author expert status; M = .65, SD = .74), and, finally, by score point 4 (evaluating the reliability of a website; M = .57, SD = .72). Thus, students' scored significantly higher on score point 1 (determining the author of the website) than on score points 2, 3, and 4. Similarly, scores on score point 3 (author's point of view) were significantly higher for students than on score point 4 (evaluating the reliability of a website). However, score point 2 (author expertise) and score point 4 (evaluating the reliability of a website) were not significantly different, meaning students performed at a similar level on these two different questions.

To address the third research question, three repeated measures ANOVAs were used to compare mean differences in CE to mean differences in each of the other four LESC skill areas (Locate, Synthesize and Communicate), within each of the three formats (Closed, Open, and Multiple Choice). The means and standard deviations of these analyses are presented in Table 2. The first repeated measures ANOVA was conducted to compare scores in the Closed format. Multivariate results show that there was a significant overall effect for LESC Skill Area in the Closed format, Wilks' Lambda = .163, F (3, 191) = 327.44, $p < .0005$, multivariate partial eta squared = .84.   Follow up, post hoc analyses of pairwise comparisons showed that each LESC Skill Area was significantly different from each other LESC skill area ($p < .005$), except for Evaluate and Communicate. This indicated that student performance in these two skill areas was not statistically different in the Closed format. Students scored higher on synthesize (M = 6.32,

SD = 1.76) than on Locate (M = 3.85, SD = 2.27), followed by Communicate (M = 3.12, SD = 1.86) and Evaluate (M = 2.84, SD = 1.54).

Table 3

Student Performance by Critical Evaluation Score Point Dimension in All Three Formats Combined

Score Point 1	Score Point 2	Score Point 3	Score Point 4	Statistical Test	Effect Size
Determining the author of the website	*Evaluating the author's expertise*	*Identifying the author's point of view and one piece of evidence that supports that point of view*	*Evaluating the overall reliability of the site using one piece of evidence from the site*	---	---
1.62 (.61)	.65 (.74)	.77 (.77)	.57 (.72)	F (3, 588) = 306.95	n_p^2 = .61

Note. p < .05.

The second one-way repeated measures ANOVA was conducted to compare scores in each of the four skill areas in the Open format. These means and standard deviations are also presented in Table 2. There was a significant effect for LESC Skill Area, Wilks' Lambda = .210, F (3, 167) = 209.14, $p < .0005$, multivariate partial eta squared = .79. Follow-up post hoc analyses of pairwise comparisons showed that each LESC Skill Area was significantly different from each other LESC skill area ($p < .005$), except Evaluate and Communicate, as was found in the Closed format. Synthesize (M = 6.10, SD = 1.74) scores averaged higher than Locate (M = 4.44, SD = 2.32), Communicate (M = 3.00, SD = 1.74), and Evaluate (M = 2.71, SD = 1.43) scores, with Locate scores ranking second highest. Communicate and Evaluate score averages were lowest.

The third one-way repeated measures ANOVA was conducted to compare scores in each of the four skill areas in the Multiple Choice format. The means and standard deviations are presented in Table 2. There was a significant effect for LESC Skill Area, Wilks' Lambda = .63, F (3, 224) = 43.19, p < .0005), multivariate partial eta squared = .37. Additionally, post hoc analyses of pairwise comparisons showed that there was significant difference (p < 0005) between Locate and Synthesize, Locate and Communicate, Evaluate and Synthesize, and Evaluate and Communicate. Evaluate scores were significantly lower (M = 4.95, SD = 1.67) than both Synthesize (M = 5.87, SD = 1.78) and Communicate scores (M = 6.10, SD = 1.67), but not significantly lower than Locate (M = 5.15, SD = 1.87) scores.

To answer the fourth and final research question, a one-way, between groups ANOVA was conducted to evaluate whether there was a significant mean difference in CE scores between the three formats, including Closed, Open, and Multiple Choice. A one-way between groups ANOVA was conducted to determine how well CE performed in each of the three LESC formats. Means and standard deviations are presented in Table 4. There was a statistically significant difference at the p < .0005 level in CE scores for the three formats: F (2, 588) = 135.69, p = .000. The effect size, measured using eta squared, was .316. Post-hoc comparisons indicated that the mean score for CE in the Multiple Choice format (M = 4.95, SD = 1.67) was significantly different from mean scores of CE in both the Closed format (M = 2.84, SD = 1.54) and the Open format (M = 2.71, SD = 1.43). Students scored higher on CE in the Multiple Choice Format than in either the Closed or Open formats. There was no statistically significant difference for CE between the Closed and Open formats.

Table 4

Student Performance on Critical Evaluation in Each of the Three Formats: Closed, Open, and

Multiple Choice M (SD)

Format Group	Format Group	Statistical Test	p
Multiple Choice 4.95	Closed 2.84 (1.54)	F (2, 588) = 135.69	.000
(1.67)	Open 2.71 (1.43)		.000

Note. Effect size: Eta squared = .316

Discussion

This study sought to determine how well seventh graders in a large, representative state sample (n = 591) critically evaluated online information. Specifically, this study examined students' performance in overall CE compared to their performance in three other skill areas, both within and across three different assessment formats. It also evaluated how well students performed in CE in each of the three formats.

Comparing CE to Locate, Synthesize and Communicate in All Formats Combined

Results from the analysis of our first research question indicated that CE was the most difficult of the four skill areas for students in all three formats combined, though the difference between CE and Communicate in the Open and Closed formats was not statistically significant. This finding supports an existing body of research that shows online CE is one of the most difficult online reading comprehension skills. As with studies of CE among older, college-aged students (Bråten, Strømsø, & Britt, 2009; Goldman, et al., 2012; Sanchez, Wiley, & Goldman, 2006), in this study, CE persisted as one of the most difficult skill areas for this younger, seventh-grade student population. The current study also demonstrates that even within

a performance-based environment like the Closed and Open formats that more closely mimics an authentic Internet context, CE was one of the most challenging of the four skill areas for students. Therefore, CE is one of the five skill areas of the new literacies of online research and comprehension (Coiro, 2003; Leu, et al., 2011) that may warrant the most instructional attention.

However, additional research is needed to determine in what ways CE is more difficult than other online reading and research skills, and how teachers should approach instruction of these skills. We do not know, for example, the types of challenges CE poses for students, or the ways in which students typically understand CE and attempt to use it when gathering sources. A follow-up qualitative analysis of students' responses to the four critical evaluation questions would be useful in adding to our understanding of this issue. Nevertheless, findings from the present study can inform both research and practice by helping to make us more aware of the significant difficulty students face when attempting to evaluate the information they find online.

Comparing the Four Dimensions of CE in All Formats Combined

Findings from the analysis of our second research question also can inform research and practice by showing us which online CE dimensions are most difficult for students and where there is a greater need to focus instruction. Students scored highest on score point one, identifying the author of the website (M = 1.62, SD =.61). This was followed in order of difficulty by score point three, or identifying the point of view of the author and a piece of evidence that supports that point of view (M = .77, SD = .77). There was no statistically significant difference in student performance between score points two and four, though score point two, evaluating the expertise of the author, had a higher mean score (M = .65, SD = .74) than score point four, evaluating the overall reliability of a site (M = .57, SD = .72).

These results show that score point two (determining author expert status) may have been more difficult for students than was score point three (providing author's point of view). One reason for this could be that score point two measures a higher-level skill than score point three. Although the score points were designed to be increasingly challenging, it appears that students actually had more difficulty determining expert status, even though it came before evaluating point of view in the task. However, it may be useful for students to evaluate the author's expertise prior to examining the author's point of view. Students' knowledge of author background and expertise may help to inform their evaluation of an author's point of view. This raises questions about whether skills in an assessment of online comprehension and research should be ordered from lower to higher levels of difficulty, or if it is more important for the questions to follow the logical sequence of the task. It may also be that an assessment that mirrors the complexities of an authentic online research experience, one in which students are naturally and logically moving back and forth between lower- and higher-level skills, is the best kind of assessment to determine students' actual capabilities.

Comparing CE to Locate, Synthesize, and Communicate in Each Format

When we investigated our third research question, we found a significant difference in the mean scores of Evaluate compared to the mean scores of Synthesize, in each of the three formats, with students scoring higher on Synthesize items than on Evaluate items. In the Closed format, the mean scores for Evaluate also were significantly lower than those for Locate. In the Multiple Choice format, the mean scores for Evaluate were significantly lower than those for Communicate. As the analyses that combined the three formats showed, the difficulty of CE persisted when we looked at its effect in each of the three formats separately, especially

compared to Synthesize. Thus, CE was one of the most difficult of the four skill areas regardless of the format in which it was assessed.

In the Closed and Open formats, it may be that Communicate posed as great a challenge as Evaluate, since students had to know how to use the email or wiki communication tools in order to be successful. In the Multiple Choice format, these questions were simplified, as students did not have to perform these actions but simply had to choose from a set of answers. Thus, it makes sense that Evaluate would be significantly harder than Communicate in the Multiple Choice format.

That CE persisted as one of the most difficult of the four skills across all three formats may suggest that all three formats are valid ways of measuring students' ability to critically evaluate information online. It may also suggest that CE is, in fact, one of the most difficult of the four online reading and research skill areas for seventh-graders, since students consistently scored lower in this skill area regardless of the format in which it was measured. Teachers should thus pay particular attention to both instruction and assessment of this important yet challenging skill.

Comparing CE in Three Formats: Closed, Open, and Multiple Choice

When we compared CE in the three formats to investigate our fourth and final research question, we found a significant difference in the mean scores of CE in the Multiple Choice format compared to both the Closed and Open formats, though there was no significant difference in mean scores between the Closed and Open formats. While the three formats were developed to be similar to one another, these results show that CE poses less of a challenge in the Multiple Choice format than it does in the other two formats.

One reason for this may be that the Multiple Choice format offers a time advantage to test takers that the other two formats do not. The four CE score points appear in a linear sequence about three quarters of the way into each assessment in all three formats. Students tended to finish the Multiple Choice sessions much more quickly than they finished the Closed or Open sessions. Thus, it is possible that students taking the Closed and Open formats were fatigued by the time they engaged in the four CE skills, while students taking the Multiple Choice format were not.

It is also important to consider that students taking the Multiple Choice test may have had a navigational advantage that students taking the other two formats did not. Students taking the Closed and Open formats had to click on a link in order to navigate to the website they were to evaluate. The CE website contained a hyperlink that students could click on in order to obtain information about the author on the author biography page. However, the student had to decide whether or not to access this page and to figure out how to access the page with additional information. In the Multiple Choice format, however, both the CE website and corresponding author biography page were presented to students alongside the question and answer choices. Thus, students taking the Multiple Choice format had a greater chance of reading both pages since they were guided to do so.

A third possible reason that CE performed better in the Multiple Choice format than in either the Closed or Open format may be that CE was measured somewhat differently in the Multiple Choice format. Because of the nature of multiple choice testing, it is possible that the presentation of the CE items may have been less complex in this format, and may therefore have required less cognitive demand for students than it did in the other two formats. Rather than generating a response to the four CE questions on their own, as they were required to do in the

Closed and Open formats, students taking the Multiple Choice assessment only had to choose from four possible answers. Each question also was presented on its own with its own images to use as reference points, whereas in the Closed, students had to manage multiple windows and types of information, including a notepad, a search window, the social networking site, and the email or wiki window. The CE task was thus much more complex in the Closed and Open formats than in the MC format.

Non-performance based assessments, such as the Multiple Choice format used in the present study, may overestimate students' critical evaluation abilities. While performance-based assessments such as the Closed and Open formats used in the current study may be more difficult and time consuming to construct and score than non-performance based formats, they may also more accurately estimate students' abilities. Test creators of multiple choice assessments, and those using and interpreting test results, should therefore keep this in mind when examining test data and forming conclusions about students' ability to critically evaluate online information.

Implications and Limitations

Findings from this study contribute to literacy research and teaching practices in several key ways. First, findings add to existing research on CE by expanding our knowledge of how students perform in CE when it is assessed in performance-based and non-performance based ways. This study is one of the first to evaluate adolescents' use of CE in an online environment within a performance-based assessment. Thus, the findings from this study, especially those that compare the three formats, are particularly informative for understanding how students actually conduct research in an online context.

Second, findings contribute to a growing body of research on CE showing it is a difficult skill area for students. CE may be one of the five online reading comprehension skills that is the

most difficult for students and thus warrants the most careful instructional attention. Findings can inform existing literature on how students perform in online CE to support future studies and practice. Findings thus inform thinking about on which online skill areas teachers should focus the most, given what many students currently are able to do. Additionally, results show with which dimensions of CE students struggle the most and thus can guide teachers to focus on teaching and assessing the most complex and nuanced skills involved in the already complex skill area of CE. This may be especially timely, as teachers will need to teach and assess these types of skills with the implementation of the new Common Core State Standards (2012) in 2014.

Finally, findings from this study raise important questions about how best to teach and assess the CE of online information. The analyses conducted in this study show that CE may be one of the most persistently difficult skills for students when reading and conducting research online. The analyses conducted in this study do not show what effective instruction that addresses deficits in CE skills might entail, and spending more time teaching CE skills will not necessarily result in increasing students' ability to effectively evaluate online information. Additionally, teachers may not have adequate technological skills to begin teaching online CE to their students. Thus, more research needs to be conducted to determine what effective versus ineffective instruction in CE of online information entails and how teachers can prepare for this instruction.

Without knowing how to teach and assess CE, we risk students learning only lower-level digital literacies skills, such as locating information, without also learning the higher-level skills necessary for using that information effectively. As teachers begin to plan for and implement the Common Core State Standards, an important question for both researchers and practitioners to ask is: What is the best approach to teaching and assessing online CE skills, which may be the

most difficult and yet also the most critical for students to learn when reading and conducting

research online?

References

Australian Curriculum Assessment and Reporting Authority. (n.d.). *The Australian Curriculum,*
  v1.2. Retrieved from www.australiancurriculum.edu.au/Home

Bennett, S., Maton, K., and Kervin, L. (2008). The 'digital natives' debate: A critical review of
  the evidence. British Journal of Educational Technology, Vol. 39, No. 5., pps. 775-786.

Bilal, D. (2000). Children's use of the Yahooligans! Web search engine: Cognitive, physical,
  and affective behaviors on fact-based search tasks. *Journal of the American Society for*
  Information Science, *51*, 646–665.

Braten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: examining the role of source
  evaluation in students' construction of meaning within and across multiple texts. Reading
  Research Quarterly, 44, 6-28.

Coiro, J. (2003). Reading comprehension on the Internet: Expanding our understanding of
  reading comprehension to encompass new literacies. *The Reading Teacher*, *56*, 458-464.

Coiro, J., Knoble, M., Lankshear, C., & Leu, D. J., Jr. (Eds.) (2008). *Handbook of research on*
  new literacies. Lawrence Erlbaum, Mahwah, NJ.

Common Core State Standards Initiative (2012). *Common Core State Standards Initiative:*
Preparing America's Students for College and Career. Retrieved from
http://www.corestandards.org.

Cope, B., & Kalantzis, M. (2000). *Multiliteracies: Literacy learning and the design of social*
  futures. London: Routledge.

Fabos, B. (2008). The price of information: Critical literacy, education, and today's Internet. In J.
  Coiro, M. Knobel, D. Leu, & C. Lankshear (Eds.). *Handbook of research on new*
  literacies (pp. 839-870). Mahwah, NJ: Erlbaum.

Goldman, S., Braasch, J., Wiley, J., Graesser, A., & Brodowinska, K. (2012). Comprehending

and learning from Internet sources: Processing patterns of better and poorer learners.

Reading Research Quarterly, 47, 356-381. doi: 10.1002/RRQ.027

Goldman, S. R., Wiley, J., & Graesser, A. C. (2005, April). Literacy in a knowledge society:

constructing meaning from multiple sources of information. Paper presented at American

Educational Research Association. Montreal, Canada

Guinee, K., Eagleton, M. B., & Hall, T. E. (2003). Adolescents' Internet search strategies:

Drawing upon familiar cognitive paradigms when accessing electronic information

sources. *Journal of Educational Computing Research*, *29*, 363–374.

Greenhow, C., Robelia, B., & Hughes, J. (2009). Web 2.0 and classroom research: What path

should we take now? *Educational Researcher, 38*(4), 246–259.

Grimes, D. & Boening, C. (2001). Worries with the Web: A look at student use of Web

resources. *College & Research Libraries*, *62(1)* 11-22.

Jenkins, H. (2006). *Convergence Culture: Where Old and New Media Collide*. New York: New

York University Press.

Judd, V.C., Farrow, L.I. and Tims, B.J. 2006. Evaluating public web site information: a process

and an instrument. Reference Service Review, Vol. 34, No. 1, pp. 12-32.

Kiili, C., Laurinen, L., & Marttunen, M. (2008). Students evaluating Internet sources: From

versatile evaluators to uncritical readers. Journal of Educational Computing Research,

39(1), 75–95.

Kress, G. (2003). *Literacy in the new media age.* London, UK: Routledge.

Kuiper, E., & Volman, M. (2008). The Web as a source of information for students in K–12

    education. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research*

    on new literacies (pp. 241–246). Mahwah, NJ: Lawrence Erlbaum.

Lankshear, C., & Knobel, M. (2006). *New literacies* (2nd ed.). Maidenhead, UK: Open

    University Press.

Leu, D. J., O'Byrne, W. I., Everett-Cacopardo, H., McVerry, J. G., Zawilinski, L.. Kiili, C.,

    Kennedy, C., Forzani, E., (2011, September). The New Literacies of Online Reading

    Comprehension: Expanding the Literacy and Learning Curriculum. Journal of Adolescent

    & Adult Literacy, 55(1), 5–14. doi: 10.1598/JAAL.55.1.1

Leu, D. J., O'Byrne, W. I., Zawilinski, L., McVerry, J. G., & Everett-Cacopardo, H. (2009).

    Expanding the new literacies conversation. *Educational Researcher, 38,* 264-269.

Leu, D.J., Jr., Kinzer, C.K., Coiro, J., Cammack, D. (2004). Toward a theory of new literacies

    emerging from the Internet and other information and communication technologies. In

    R.B. Ruddell & N. Unrau (Eds.), *Theoretical Models and Processes of Reading*, *Fifth*

    Edition (1568-1611). International Reading Association: Newark, DE.

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., Henry, L. A. (2013). New literacies: A dual level

    theory of the changing nature of literacy, instruction, and assessment. In Alvermann, D.

    E., Unrau, N. J., & Ruddell, R. B. (Eds.), *Theoretical Models and Processes of Reading*,

    Sixth Edition. Newark, DE: International Reading Association.

Lewis, C., & Fabos, B. (2005). Instant messaging, literacies, and social identities. *Reading*

    Research Quarterly, *40*, 470–501.

Organisation for Economic Co-operation and Development & the Centre for Educational Research and Innovation. (2010). *Trends shaping education 2010.* Paris: OECD Publishing.

Prensky, M. (October, 2001). Digital natives, digital immigrants. *On the Horizon*. 9 (5). MCB University Press. Available at:

http://www.albertomattiacci.it/docs/did/Digital_Natives_Digital_Immigrants.pdf

Rieh, S. Y., & Becklin, N. J. (1998). Understanding judgment of information quality and cognitive authority in WWW. *Journal of the American Society for Information Science*, *35*, 279-289.

Sanchez, C. A., Wiley, J., & Goldman, S. R. (2006). Teaching students to evaluate source reliability during Internet research tasks. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Proceedings of the seventh international conference on the learning sciences* (pp. 662-666). Bloomington, IN: International Society of the Learning Sciences.

Sutherland-Smith, W. (2002). Weaving the literacy Web: Changes in reading from page to screen. *The Reading Teacher, 55,* 662–669.

Wallace, R. M., Kupperman, J., Krajcik. J., & Soloway, E. (2000). Science on the Web: Students Online in a Sixth-Grade Classroom. *Journal of the Learning Sciences, 9,* 75-104.

Wiley, J., Goldman, S., Graesser, A., Sanchez, C., Ash, I., & Hemmerich, J. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal, 46,* 1060–1106. doi:10.3102/0002831209333183