How Well Do Students Judge Their Ability to Critically Evaluate Online Information?: Results from an Internet Use Survey and an Online Reading Comprehension Assessment

Elena Forzani, Cheryl Maykel, and Clint Kennedy, University of Connecticut

Abstract

This study investigated how well students in a stratified, random, two-state sample evaluated their ability to critically evaluate online information in two different assessment formats. The formats included a Closed, or simulated, Internet environment (n = 1,409) and a Multiple Choice (n = 1,589) environment. Results indicated that overall, students were not effective at judging their ability to critically evaluate online information. While none of the four dimensions of Critical Evaluation (CE) that were assessed contributed a large percentage of variance to actual, overall, total CE score, students' evaluation of their ability to evaluate author point of view contributed the most. Findings suggest that teachers may need to use instructional methods for teaching CE that help students better judge their abilities and thus be more likely to invest themselves in the learning process for online CE.

How Well Do Students Judge Their Ability to Critically Evaluate Online Information?: Results from an Internet Use Survey and an Online Reading Comprehension Assessment

The Internet and other information and communication technologies (ICTs) have significantly altered the ways in which we view and use literacy skills (Coiro, Knobel, Lankshear, & Leu, 2008). The increasing importance of the Internet, especially as a fundamental tool in our everyday and working lives, has necessitated both new literacy skills as well as new ways of thinking about traditional literacy skills necessary for proficiency on the Internet.

One of the most significant changes in literacy as a result of the Internet has been the degree to which the critical evaluation (CE) of information influences our ability to comprehend and synthesize the texts we locate and read online. Indeed, CE has become one of the most important skill sets required by readers today (Goldman, et al., 2012; Wiley, Goldman, Graesser, Sanchez, Ash, Hemmerich, 2009). Yet, it is often one of the most difficult areas for students in online research and comprehension (Forzani & Burlingame, 2012; Kuiper & Volman, 2008).

Students who struggle with CE may not be effective at judging their abilities in this area, as higher-performing students have been shown to be better at predicting their academic abilities than middle and lower-performing students (Cole & Gonyea, 2010; Kuncel, Crede & Thomas, 2005). This is problematic, since those students who most need CE skills may not understand the importance of learning these skills, and thus may put less time and effort into instructional activities in these areas. Additionally, teachers may be reluctant to teach technology-related skills because they believe students are already proficient in this area since they are "digital natives" (Prensky, 2001).

The present study thus investigated how well students were able to judge their ability to critically evaluate online information. The purpose of this study was to investigate the extent to

which students' self-report of critical evaluation ability predicted their actual CE scores on the ORCA (Online Reading Comprehension Assessment).

Perspectives and Theoretical Background

This study is framed by several perspectives useful for thinking about students' ability to judge their skill at using online CE skills. These perspectives include a dual level theory of New Literacies (Coiro, Knobel, Lanskshear, & Leu, 2008; Leu, O'Byrne, Zawilinski, McVerry, & Everett-Cacopardo, 2009), perspectives on the critical evaluation of information, and perspectives on student self-report of academic performance.

New Literacies: A Dual Level Theory

A dual-level theory of New Literacies (Leu, O'Byrne, Zawilinski, McVerry, & Everett-Cacopardo, 2009; Leu, Kinzer, Coiro, Castek, & Henry, 2013) conceives of literacy on two levels: uppercase (New Literacies) and lowercase (new literacies). *New Literacies*, which is the broader concept, encompasses the more rapidly changing, lowercase, dimensions of *new literacies.* Lowercase theories explore a specific area of new literacies (Lewis & Fabos, 2005), or a particular discipline (Kress, 2003). Common findings across multiple perspectives can be integrated into a broader New Literacies theory that is likely to be more stable over time as a result.

Within the dual-level theory of New Literacies, online reading often occurs as part of a research task. At least five processing practices occur during online research and comprehension that include both traditional and new skills and strategies: 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., in press; Jenkins, 2006); and 5) reading and writing to communicate information (Greenhow, Robelia, & Hughes, 2009).

Many areas of New Literacies research view the Internet as a tool that allows for the use of new types of online social practices (Greenhow, Robelia, & Hughes, 2009; Lankshear & Knobel, 2006). In the present study, both formats of the assessment were situated within a social network that required students to interact via instant messages, wikis, and emails while conducting and reporting on their research. The four skill areas of locating, evaluating, synthesizing, and communicating were assessed within the context of an online research task. We focused particular attention in this study on the critical evaluation of online information. Specifically, we focused on the evaluation of author, point of view, and the reliability of source. Students therefore used many lowercase new literacies throughout the assessment that were themselves situated within the broader uppercase New Literacies.

Critical Evaluation

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

Much previous research on critical evaluation within an online context has focused on college students' abilities (Bråten, Strømsø, & Britt, 2009; Goldman, et al., 2012; Sanchez, Wiley, & Goldman, 2006). This research has led to the inclusion of critical evaluation and higher-level thinking in the recent Common Core State Standards (2012) in the U.S. This research also has impacted curricular frameworks for K-12 education in other nations such as the recent Australian Curriculum (Australian Curriculum Assessment and Reporting Authority, n.d.). However, while this has allowed us to better understand college-aged students' ability to evaluate information, especially online information, we know much less about younger students' ability to critically evaluate online sources. Given this is now part of many nations' curricular frameworks, as well as part of the Common Core standards, 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 their curriculums.

One important aspect of understanding students' ability to critically evaluate online information is the tendency for students to overestimate their abilities in this area (Grimes & Boening, 2001). The combination of the ease with which students are able to locate information, along with their lack of awareness of the difficulty involved in evaluating information, could easily lead to students collecting, reporting, learning, and using invalid information to guide their research. The implications of students' potential lack of awareness of the difficulty involved in evaluating online information on their future growth and scholarly activity are, therefore, enormous.

Self-Report

In addition to better understanding how well younger children can critically evaluate online information, it is also useful to understand how students view themselves as critical

evaluators, since this impacts instructional approaches. It may be that if students feel they are already skilled in this area, they are likely to be less invested in learning these skills. However, if students become aware of their own need for these skills, they may be more likely to be motivated to learn these skills and less likely to knowingly misuse potentially inaccurate information.

Research on students' self-report of general academic performance shows that high-performing students are more likely to accurately report their academic scores than are their low-performing peers (Cole & Gonyea, 2010; Kuncel, Crede & Thomas, 2005). Therefore, we expected to find that students who scored higher on the CE items would also report that they were more skilled in CE. However, we do not yet know if students will report on their ability in CE skills similarly to how students in the previous work reported on their general academic ability.

Thus, this study sought to determine how well students gauged their performance in four different aspects of critical evaluation: determining the author of a website, evaluating the author's expert status, identifying the author's point of view and evidence from the website to support that point of view, and evaluating the overall reliability of a website. We wanted to know whether students' evaluations of their skills on the Internet Use Survey (IUS) significantly predicted their CE scores on the ORCA in two different assessment formats. We used regression analyses to analyze these relationships in both a performance-based, closed Internet assessment context (ORCA-Closed, a simulated Internet environment), and in a multiple choice context (ORCA-Multiple Choice). Specifically, we investigated the following three research questions: 1) How well did students' overall IUS CE scores (a composite of their four individual CE-related IUS scores) predict their actual total CE score on the ORCA in two different assessment

formats?; 2) How well did students' scores on the four CE-related questions of the IUS, that directly corresponded to four CE score points on the ORCA, predict students' actual performance on these items of the ORCA, in each of the two different assessment formats?; 3) How much unique and significant variance did each of the four IUS score points contribute to the actual, overall, total CE score on the ORCA?

Method

Participants

Representative samples of seventh grade students from each of two states in the northeastern United States were used. Participants for the investigation of the Closed version of the assessment included a total of 1,409 students from these two states, with 748 students from State 1 and 661 students from State 2. Participants from the MC assessment included a total number of 1,589 students from both states, with 870 students from State 1 and 719 students from State 2. Students who participated in an assessment in one format also participated in an assessment in the other format except in cases of data loss (see below for a more complete description of this).

Participants included students from 17 school districts in State 1 and 23 districts in State 2, for a total of 40 districts, with one participating school per district. For the majority of the schools, one teacher with two classes of approximately 25 students each participated except in the case of a few smaller schools where it was necessary to include more than one teacher. Both districts and schools from each state were selected using stratified random sampling. The sampling plan stratified schools within each state, while taking note of school size, according to three factors: 1) school 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). Schools were randomly sampled within each of these strata and invited to participate.

Principals at each of the selected schools identified an English Language Arts teacher or teachers (in the case of smaller schools) whose students best represented the entire school population and who were willing to participate. Teachers then selected two of their classes that best fit this same description. Students from the selected classrooms who wished to participate provided a signed parental consent form and a completed student assent form. The majority of students completed both of the assessments. However, due to absences, system crashes, and data capture glitches, 17% percent of the overall sample did not complete the assessment, and 6.5% percent of the overall sample did not complete the MC assessment. Thus, the differences in the numbers of students in the Closed and MC assessments are due to data loss.

Online Research and Comprehension Assessments (ORCAs)

Four research scenarios, using four different life science topics, were developed as part of The ORCA Project (Leu, Kulikowich, Sedransk, & Coiro, 2009). All of the assessments required students to read and conduct research using the Internet. These assessments have previously demonstrated high levels of reliability and validity (Leu, Coiro, Kulikowich, & Cui, 2012). Each of these four scenarios was developed in two different assessment formats. These formats included a Closed Internet environment and a Multiple Choice environment (ORCA-Closed and ORCA-MC).

The ORCA-Closed consisted of a simulated Internet environment that was online, but also closed off to the full, actual Internet. This environment required students to search for, select, and use websites from the project's own search engine, called "Gloogle," which was

populated with a predetermined set of websites. The Closed format was thus largely a performance-based measure.

The ORCA-MC, on the other hand, confined students to selecting sites and answers from a set of four answer choices per question, though it was modeled after the ORCA-Closed. 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 answer each question successfully. Students could toggle between the different screenshots as needed. Thus, the ORCA-MC attempted to provide students with a richer context than some more traditional multiple-choice assessments, while maintaining a more efficient, multiple-choice structure.

In all scenarios, students were presented with science research problems focused on health and human body systems, since this is a topic common to many seventh grade science curricula. Two of the scenarios focused on the eyes, and two of them focused on the heart. The topics included energy drinks and heart health, video games and eye health, snacks and heart health, and cosmetic contact lenses and eye health. 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 form a main idea about what they learned from their research. ICC scenarios asked students to investigate both sides of the issue and take a position.

Each scenario included items assessing students' ability to locate, evaluate, and synthesize information found during the research process, as well as items assessing students' ability to communicate the results of their research via email or wiki. Each scenario, called a

LESC, represented each of the four skills areas of Locate, Evaluate, Synthesize, and Communicate. Each of these skill areas included three process skills and one product skill, with one score point assessing each skill. This resulted in a total of four score points for each of the four skill areas, and a total of 16 score points for each LESC.

The LESC questions appeared within a Facebook-like environment involving 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 through the research process, and invited to engage in the four skill areas through various prompts and questions from Brianna and Jordan.

The four CE score points (see Table 1) were directly related to three of the traditional critical evaluation criteria that included 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). 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 link he provided. From that website, students had the opportunity to navigate to the author biography page, which was hyperlinked to the given site and allowed them to gather more information about 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 significantly. Students were asked to determine the author of a given website (authority), evaluate the author's

expertise (authority), identify the author's point of view and a supporting detail (objectivity), and evaluate the overall reliability of the site using at least one piece of valid reasoning (accuracy).

Student Internet Use Survey

In addition to the ORCA measure, students completed an Internet Use Survey (IUS) on Survey Monkey. Responses were recoded from the Survey Monkey output into a database that housed all data for the project. The four questions analyzed here were directly related to each of the four critical evaluation items (see Table 1). In other words, students were asked to evaluate their ability to engage in each of the four critical evaluation skills that were assessed by the ORCA.

Students' responses were recoded from four initial categories to create a dichotomous variable so that answer choices included either, 1) I don't know how to do this *and* My skills are limited in this area, or, 2) My skills are average in this area *and* My skills are excellent in this area. This variable thus dichotomized students' responses into two groups with the first group consisting of students who did not feel they were able to engage in this skill well and the second group consisting of students who felt they were able to engage in this skill relatively well.

Table 1

CE Items	Corresponding IUS Items
1. Can the student identify the author of a website?	1. I can find out who the author is of a website.
2. Can the student evaluate the author's level of expertise?	2. I can determine if a website's author is an expert or not.
3. Can the student identify the author's point of view?	3. I can determine the point of view of a website's author.
4. Can the student evaluate the reliability of a website?	4. I can evaluate the reliability of information at a website.

Procedures

LESC Administration.

The ORCAs were administered in two separate sessions on each of two assessment days at each school. On the first assessment day, the test administrator read students brief, standardized instructions before beginning the assessment. The ORCA used an automated start-up sequence on a set of MacBook Airs. By entering their unique identification numbers into the login screen, students were brought directly to their assigned ORCA in the online system, on each of the two assessment days. Students received the same accommodations during assessments that they typically received in the classroom.

An auto-capture system recorded students' online reading performance for both product and process score points for later scoring. Four score points were calculated for each of the four major skill areas (Locate, Evaluate, Synthesize, and Communicate) using a binary (1 or 0) score point system within each LESC. Therefore, students had the opportunity to earn a total of 16 score points in each LESC.

The ORCA scoring system automatically scored the ORCA-MC assessments, but the ORCA-Closed reports were hand-scored by a team of four undergraduate scorers, one for each of the four scenarios. Each scorer was trained to a minimum inter-rater reliability level of 90% accuracy for each score point. Scorers were then required to meet 90% accuracy when checked against an expert scorer for roughly 20% of each of the score points. Scorers who did not meet 90% accuracy were retrained and retested to this level before continuing scoring.

Analysis.

Standard multiple regression analysis was used to answer all three of the study's research questions: 1) How well does a student's overall survey score in four dimensions of CE combined

predict their actual performance on the total CE score for the ORCA, for each of the two assessment formats?; 2) How well does student self-report of ability to critically evaluate online information in four dimensions predict actual, ability to critically evaluate online information in those four dimensions, for each of two formats?; and 3) How much separate and independent variance does each of the four predictors contribute to the actual overall, total CE score, in each of the two formats?

Results

Prior to the statistical analyses, all data were examined and found to meet assumptions of regression analysis, including linearity, homoscedasticity, and multivariate normality. The means, standard deviations, ranges, and inter-correlations for all variables in the different analyses are presented in Tables 2 through 6.

Table 2

Research Question 1 (Closed and MC): Means, Standard Deviations, and Intercorrelations Among Variables

Reader Variable	IUSQSUM (Closed)	IUSQSUM (MC)
1. Actual, overall CE Total (Closed) (M = 1.49; SD =.980)	.140***	
2. Actual, overall CE Total (MC) (M = 2.70; SD = 1.14)		.162***
N	1409	1589
M	6.53	6.56
SD	3.22	3.20

CE = Critical Evaluation. ^{ns}Not significant. *** p <.001.

Table 3

Research Question 2 (Closed) Means, Standard Deviations, and Inter-correlations Among Variables

Reader Variable	IUSQ51	IUSQ52	IUSQ53	IUSQ54
1. EVA1/Author Score Point (*M* = .82; *SD* = .382)	.070**			
2. EVA2 Score Point/Author Expert (*M* = .22; *SD* = .414)		.064*		
3. EVA3 Score Point/POV (*M* = .30; *SD* = .460)			.131***	
4. EVA4 Score Point/Reliability (*M* = .14; *SD* = .350)				.100***
M	.687	.534	.607	.632
SD	.464	.499	.489	.483

CE = Critical Evaluation. *N* = 1,336 for all Survey Question (independent) variables and 1,409 for all Score Point (dependent) variables. *p <.05. ** p <.005. *** p <.001

Table 4

Research Question 2 (MC) Means, Standard Deviations, and Intercorrelations Among Variables

Reader Variable	IUSQ51	IUSQ52	IUSQ53	IUSQ54
1. EVA1 Score Point (*M* = .73; *SD* = .442)	.121***			
2. EVA2 Score Point (*M* = .68; *SD* = .465)		.079***		
3. EVA3 Score Point (*M* = .66; *SD* = .486)			.096***	
4. EVA4 Score Point (*M* = .62; *SD* = .485)				.086***
M	.692	.536	.617	.642
SD	.462	.499	.486	.480

CE = Critical Evaluation. *N* = 1,502 for all Survey Question (independent) variables and 1,589 for all Score Point (dependent) variables. *** $p \leq .001$

Table 5

Research Question 3 (Closed): Means, Standard Deviations, and Intercorrelations Among

Reader Variable	1	2	3	4	5
1. Actual, overall CE Total	1.000				
2. Self-report of first CE dimension	.084	1.000			
3. Self-report of second CE dimension	.101	.469	1.000		
4. Self-report of third CE dimension	.160	.396	.477	1.000	
5. Self-report of fourth CE dimension	.121	.409	.512	.482	1.000
M	1.49	.687	.534	.607	.632
SD	.980	.464	.499	.489	.483

CE = Critical Evaluation. $N = 1{,}336$ across all variables. $p \leq .001$

Table 6

Research Question 3 (MC): Means, Standard Deviations, and Intercorrelations Among

Reader Variable	1	2	3	4	5
1. Actual, overall CE Total	1.000				
2. Self-report of first CE dimension	.139	1.000			
3. Self-report of second CE dimension	.112	.450	1.000		
4. Self-report of third CE dimension	.182	.401	.462	1.000	
5. Self-report of fourth CE dimension	.116	.390	.508	.454	1.000
M	2.70	.692	.536	.617	.642
SD	1.140	.462	.499	.486	.480

CE = Critical Evaluation. $N = 1{,}502$ across all variables. $p < .001$.

Research Question 1

Closed.

To investigate the first research question, two new variables were created for the Closed and MC analyses. One of these variables represented the sum of the four IUS questions addressing students' CE abilities, and the other represented the sum of the four CE score points on the ORCA. A regression analysis was conducted for each of the two assessment formats using the IUS Sum variable as the predictor, and the CE Total from the ORCA as the dependent variable. In the Closed, the R Square value was .020 (see Table 7). Thus, the sum of the four IUS question

scores predicted about 2.0% of the variance in overall, total CE score. This was significant at the p = .000 level. Additionally, there was a small but significant and positive correlation, r = .140, p = .000.

Multiple Choice.

Similarly, for the Multiple Choice, the R Square was .026 (see Table 7). Thus, the sum of the four IUS question scores predicted about 2.6% of the variance in overall, total CE score. This was significant at the p = .000 level. Additionally, there was a small but significant and positive correlation, r = .162, p = .000.

Table 7

Research Question 1 (Closed and MC): Regression Analyses of Students' Evaluation of Their Ability in Four Dimensions of Critical Evaluation Combined on Students' Actual, Overall Critical Evaluation Ability

Dependent and independent variables	*R*	R^2	Adjusted R^2	*F*	β
Actual, overall CE Total					
IUSQSUM: Self-report of all four CE dimensions combined (Closed)	.140	.020	.019	28.244***	.140
IUSQSUM: Self-report of all four CE dimensions combined (MC)	.162	.026	.026	42.6500***	.162

CE = Critical Evaluation. *** $p \leq .001$

Research Question 2

Closed.

To investigate the second research question for the Closed format, four separate regression analyses were conducted. For the Closed assessment format, for IUSQ51, which corresponded to the first dimension of CE tested on the ORCA, the R Square was .005 (see Table 8). Thus, IUSQ51 explained only about .5% of the variance in the E1 score point. This was significant at the p = .011 level. For IUSQ52, which corresponded to the second dimension of

CE tested on the ORCA, the R Square was .004. Thus, IUSQ52 explained about .4% of the variance in the E2 score point. This was significant at the p = .020 level. For IUSQ53, which corresponded to the third dimension of CE tested on the ORCA, the R Square was .017. Thus, IUSQ53 explained about 1.7% of the variance in the E3 score point. This was significant at the p = .000 level. For IUSQ54, which corresponded to the fourth dimension of CE tested on the ORCA, the R Square was .010. Thus, IUSQ54 explained about 1.0% of the variance in the E4 score point. This was significant at the p = .000 level.

Table 8

Research Question 2 (Closed): Regression Analyses of Students' Evaluation of Their Ability in Four Dimensions of Critical Evaluation on Students' Actual, Critical Evaluation Ability in Each Dimension

Independent Variables	*R*	R^2	Adjusted R^2	*F*	β
IUSQ51: Self-report of first CE dimension	.070	.005	.004	6.559**	.070
IUSQ52: Self-report of second CE dimension	.064	.004	.003	5.406**	.064
IUSQ53: Self-report of third CE dimension	.131	.017	.017	23.452***	.131
IUSQ54: Self-report of fourth CE dimension	.100	.010	.009	13.597***	.100

CE = Critical Evaluation. ** $p < .05$. *** $p \leq .001$

Multiple Choice.

To investigate the second research question with the Multiple Choice format, four additional analyses were conducted. For IUSQ51, which corresponded to the first dimension of CE tested on the ORCA, the R Square was .015 (see Table 9). Thus, IUSQ51 explained 1.5% of the variance in the E1 score point. This was significant at the p = .000 level. For IUSQ52, which corresponded to the second dimension of CE tested on the ORCA, the R Square was .006. Thus, IUSQ52 explained 6% of the variance in the E2 score point. This was significant at the p

= .002 level. For IUSQ53, which corresponded to the third dimension of CE tested on the ORCA, the R Square was .009. Thus, IUSQ53 explained .9% of the variance in the E3 score point. This was significant at the p = .000 level. For IUSQ54, which corresponded to the fourth dimension of CE tested on the ORCA, the R Square was .007. Thus, IUSQ54 explained .7% of the variance in the E4 score point. This was significant at the p = .001 level.

Table 9

Research Question 2 (MC): Regression Analyses of Students' Evaluation of Their Ability in Four Dimensions of Critical Evaluation on Students' Actual, Critical Evaluation Ability in Each Dimension

Independent Variables	*R*	R^2	Adjusted R^2	*F*	β
IUSQ51: Self-report of first CE dimension	.121	.015	.014	22.141***	.121
IUSQ52: Self-report of second CE dimension	.079	.006	.006	9.490***	.079
IUSQ53: Self-report of third CE dimension	.096	.009	.009	13.926***	.096
IUSQ54: Self-report of fourth CE dimension	.086	.007	.007	11.133***	.086

CE = Critical Evaluation. *** $p \leq .002$.

Research Question 3

Closed.

For the analyses in each of the two formats for our third research question, all five variables were entered into the model together. For the Closed format analysis, the R Square was .028 (see Table 10). Thus, all four predictors together explained only about 2.8% of the variance in actual Total CE Score. This was significant at the p = .000 level.

The beta value for students' self-evaluation of the third dimension of CE (see Table 10) was the highest, beta = .128, p = .000), showing that students' self-evaluation of the third dimension of CE was the best predictor of overall, total CE score. The part correlation is .105. Thus,

students' evaluation of the first CE dimension explained 1.05 % of the variance in overall, Total CE Score. Correlation statistics in Table 6 also show that this predictor had the highest correlation with overall, actual CE total. The third CE dimension had a small but significant and positive correlation, $r = .160$, $p < .000$, with overall, actual CE total.

The first predictor, students' self-evaluation of the first dimension of CE, had a beta value of .008, $p = .812$ (see Table 10). Correlation statistics in Table 5 show that this predictor had the lowest correlation with overall, actual CE total. The first CE dimension had a very small, positive, and significant correlation, $r = .084$, $p = .001$, with overall, actual CE total.

The second predictor, students' self-evaluation of the second dimension of CE, had a beta value of .010, $p = .768$ (see Table 10). Correlation statistics (see Table 5) also show that this predictor had the second lowest correlation with overall, actual CE total. The second CE dimension had a very small, positive, and significant correlation, $r = .101$, $p = .000$, with overall, actual CE total.

The fourth predictor, students' self-evaluation of the fourth dimension of CE, $b = .051$, was also not significant, with $p = .131$ (see Table 10). Correlation statistics (see Table 5) also show that this predictor had the second highest correlation with overall, actual CE total. The second CE dimension had a very small but significant and positive correlation, $r = .121$, $p = .000$, with overall, actual CE total.

Thus, students' evaluation of the third dimension of CE was the only predictor that made a significant, unique contribution to actual CE score. The correlations with the other predictors were significant, though they were small.

Table 10

Research Question 3 (Closed): Regression Analyses of Students' Evaluation of Their Ability in Four Dimensions of Critical Evaluation on Students' Actual, Overall Critical Evaluation Ability

Dependent and independent variables	*R*	R^2	Adjusted R^2	*F*	β
Actual, overall CE Total					
All four predictors together	.168	.028	.025	9.670*	
Self-report of first CE dimension					$.008^{ns}$
Self-report of second CE dimension					$.010^{ns}$
Self-report of third CE dimension					.128*
Self-report of fourth CE dimension					$.051^{ns}$

CE = Critical Evaluation. ^{ns}Not significant. * $p < .001$.

Multiple Choice.

For the MC analysis, the R Square was .036 (see Table 11). Thus, all four predictors together explained only about 3.6% of the variance in actual Total CE Score. This was significant at the p = .000 level. The beta value for students' self-evaluation of the third dimension of CE (see Table 11) was the highest, beta = .142, p = .000, showing that students' self-evaluation of the third dimension of CE was the best predictor of overall, total CE score. The part correlation is .118. Thus, students' evaluation of the first CE dimension explained 1.18 % of the variance in overall, Total CE Score. Correlation statistics (see Table 6) also show that this predictor had the highest correlation with overall, actual CE total. The third CE dimension had a small but significant and positive correlation, r = .182, p < .000, with overall, actual CE total.

The first predictor, students' self-evaluation of the first dimension of CE, had the second highest beta value, .072, p = .015 (see Table 11). Correlation statistics (see Table 6) show that this predictor also had the second highest correlation with overall, actual CE total. The first CE dimension had a very small, positive, and significant correlation, r = .139, p = .000, with overall, actual CE total.

The second predictor, students' self-evaluation of the second dimension of CE, had the lowest beta value, .002, p = .939 (see Table 11). Correlation statistics (see Table 6) also show that this predictor had the lowest correlation with overall, actual CE total. The second CE dimension had a very small, positive, and significant correlation, r = .112, p = .000, with overall, actual CE total.

The fourth predictor, students' self-evaluation of the fourth dimension of CE, b = .022, was also not significant, with p = .721 (see Table 11). Correlation statistics (see Table 6) also show that this predictor had the third highest correlation with overall, actual CE total. The fourth CE dimension had a very small but significant and positive correlation, r = .116, p = .000, with overall, actual CE total.

Thus, as with the Closed version, students' evaluation of the third dimension of CE was the only predictor that made a significant, unique contribution to actual CE score. While the correlations of the other predictors with overall, actual total CE score were significant, they were very small.

Table 11

Research Question 3 (MC): Regression Analyses of Students' Evaluation of Their Ability in Four Dimensions of Critical Evaluation on Students' Actual, Overall Critical Evaluation Ability

Dependent and independent variables	R	R^2	Adjusted R^2	F	β
Actual, overall CE Total					
All four predictors together	.197	.039	.036	15.117	
Self-report of first CE dimension					.072*
Self-report of second CE dimension					$.002^{ns}$
Self-report of third CE dimension					.142**
Self-report of fourth CE dimension					$.022^{ns}$

CE = Critical Evaluation. ^{ns}Not significant. *p <.05. ** $p \leq .001$.

Discussion

This study sought to determine how well student self-report of four dimensions of CE in an online reading and research task predicted both students' ability in each of those four dimensions and students' actual, overall ability to critically evaluate online information. The four dimensions included: 1) identifying the author of a webpage; 2) evaluating the author's expertise; 3) identifying the author's point of view; and 4) evaluating the reliability of the webpage. Actual, overall, ability to critically evaluate online information was a composite score that included each of the four CE dimensions given above.

Results indicated that the score point in the third dimension, students' evaluation of their ability to identify the author's point of view and a piece of evidence that matched that point of view, contributed the greatest percentage of variance to the actual, overall, CE score in both the Closed and Multiple Choice formats. Of the four CE IUS questions, the third dimension of CE also had the highest percentage of variance explained for the corresponding ORCA score point. One reason for this may be that this third dimension of CE is the most complex of the four skills. While the amount of variance accounted for is small, it is also important to consider that there are many other factors that may account for variance in actual CE score that were considered in this model, such as offline critical evaluation and prior knowledge of source material. Additional analyses could include these variables to examine their potential effects. Still, this analysis provides a useful initial look at the effect of this first predictor.

Another reason that students' evaluation of author point of view was the best of the four predictors of the total CE score may be that this skill is used with offline texts, and thus is more often taught to students. Also, this skill may not be as different in an online context as are some of the other CE skills. For example, once students are on the webpage containing the article, they

are reading and evaluating the author's point of view much like they would if reading a traditional text. However, in order to determine the author of the website and that author's expert status (CE score points 1 and 2), students have to know both how to navigate to the appropriate place on the webpage and where to click to open a second page with additional information.

Thus, it may be easier for students to understand what the skill of evaluating the author's point of view entails, since they have more experience with it. The view from the new literacies theory of online research and comprehension (Coiro, 2003; Leu, et al., 2011) is that online reading requires an interweaving of both online and offline skills, and this is what we see here. This suggests that students need more experience with online skills so they can gain a better sense of what these skills entail and how to apply what they already know from reading and researching in an offline environment. Also, more experience may help students to better judge their ability to engage in these skills. Both of these possibilities provide important information about how well each of the four IUS items can predict overall CE score.

Results also indicated that overall, students were not effective at judging their own ability to critically evaluate online information. Together, the four predictors explained only about 2.8% of the variance in overall, total CE score. This was similar to what we found using the IUS Sum variable, which contributed 2.0% of the variance in actual, total CE score for the Closed and 2.6% for the Multiple Choice. It may be that students currently in the seventh grade have grown up around new technologies, especially the Internet, and believe they are capable of engaging in any skill that requires using online information. It may also be that because students are generally less skilled at critical evaluation, they are also less skilled at evaluating their own abilities with these types of skills. This theory makes sense given the typical pattern of student self-evaluation, where higher performing students are better at accurately evaluating their

academic abilities than are lower-performing students (Cole & Gonyea, 2010; Kuncel, Crede & Thomas, 2005).

The present analysis, however, only examined critical evaluation. Additional analyses could investigate whether students are ineffective at judging their ability to locate, synthesize, and communicate information as well. This would help to determine whether online CE is an especially hard skill to for students to judge their ability in, or whether all four of these online reading and research skills pose challenges to students in terms of evaluating their abilities. Even so, the present study provides important initial findings about how well students can judge their ability to critically evaluate online information, which can contribute greatly to instruction in these skills.

Implications

Findings from this study contribute to literacy research in several key ways. First, findings add to existing research about both critical evaluation and a dual-level theory of new literacies by providing a more complete picture of students' CE performance and how it relates to self-report data. Not only is critical evaluation one of the most difficult skills for students to master (Kuiper & Volman, 2008), but it may also be a skill for which students struggle to judge their own ability. Researchers who study critical evaluation should thus pay close attention to the measures they use to evaluate students' abilities. Using self-report measures to determine students' abilities to critically evaluate online information may lead to erroneous conclusions about the data, including an overestimation of students' abilities. It is possible, then, that this is also the case for other online reading and research skill areas that are identified in theories of online reading and research, including locating, synthesizing, and communicating information. Additional research could examine how well students' self report of these skills predicts their

actual performance in these areas to determine if we see a similar pattern as we do with critical evaluation. These additional skills are important aspects of online reading and research comprehension, a lower-case new literacies skill.

Second, it is also important to consider that students may overestimate their abilities to critically evaluate online information because they spend much of their time online and thus believe they are adept at anything that occurs online. This is important for teachers, as many teachers may think that since students have grown up with many technologies, and especially with the Internet, they already know many of the skills necessary for reading and researching online. Moreover, many teachers may assume that if students are skilled in these areas, that they also are competent at judging their own abilities in these areas. However, the findings of the present study show that this is not necessarily the case. Teachers cannot always rely on students' own evaluation of their ability to critically evaluate online information. Teachers should thus be certain to pay more attention to actual assessment of student performance and not allow common assumptions about students' abilities to affect their conclusions about students' actual skill levels. Assessing students' online CE abilities can more accurately be accomplished by giving students an actual assessment rather than by using self-report data. These additional assessments will help teachers gain a more complete understanding of students' existing capabilities. Additionally, when teaching online CE, teachers should be aware that students may overestimate their abilities in this area so they can help invest students in learning these skills prior to beginning a unit of instruction, since students may not already be invested. Prior to instruction, students will need to first understand the importance of spending time on these skills.

Finally, findings from this study raise important questions about how best to teach students to critically evaluate online information. How should teachers frame instruction of

critically evaluating online information? And, how should teachers teach students who already believe they are skilled at critically evaluating online information? Teachers may be more effective if they frame instruction around initial discovery tasks to help students gain better insight into their own abilities. Without knowing how to effectively teach and assess critical evaluation in ways that will produce learning gains for students, we risk students learning only lower level digital literacy skills without learning the higher-level skills that inform them about how to use that information effectively, an important aspect of a dual-level theory of new literacies. Finally, teachers who may be reluctant to teach students online skills because they believe students are already skilled in these areas may now have more of a reason to believe that teaching these skills to students is important. These are important considerations as teachers begin to teach the Common Core standards, which require students to critically evaluate their online sources before using those sources in a report. As students gain more experience with online CE, hopefully they will become more skilled at judging their abilities in this important area.

References

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

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.

Cole, J.S. & Gonyea, R. M. (2010). Accuracy of self-reported SAT and ACT test scores: Implications for research. *Research in Higher Education, 51*, 305-319. doi: 10.1007/s11162-009-9160-9

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.

Forzani, E., & Burlingame, C. (2012, December). *Evaluating Representative State Samples of Seventh-Grade Students' Ability to Critically Evaluate Online Information*. Paper presented at the 62nd Annual Conference of the Literacy Research Association, San Diego, CA.

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

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

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.

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.

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: 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.

Kuncel, N.R., Crede, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75 (1*), 63-82.

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

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.

Leu, D. J., Kulikowich, J., Sedransk, N., & Coiro, J. (2009). Assessing online reading comprehension: The ORCA project. Research grant funded by the U.S. Department of Education, Institute of Education Sciences.

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

Lewis, C., & Fabos, B. (2005). Instant messaging, literacies, and social identities. *Reading Research Quarterly*, *40*, 470–501.

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.

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