This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Kiili, Carita; Leu, Donald J.; Utriainen, Jukka; Coiro, Julie; Kanniainen, Laura; Tolvanen, Asko; Lohvansuu, Kaisa; Leppänen, Paavo H.T.

Title: Reading to Learn From Online Information : Modeling the Factor Structure

Year: 2018

Reading to Learn from Online Information: Modeling The Factor Structure

Carita Kiili, PhD (Corresponding author)
Department of Education, University of Oslo
P.O. Box 1092, Blindern, 0137 Oslo, Norway &
Department of Education, University of Jyvaskyla
email: c.p.s.kiili@iped.uio.no

Professor Donald J. Leu
Neag School of Education, University of Connecticut
email: donald.leu@uconn.edu

Jukka Utriainen, Project Researcher, University of Jyväskylä, Finland
email: jukka.t.utriainen@jyu.fi

Associate Professor Julie Coiro, University of Rhode Island
email: jcoiro@snet.net

Laura Kanniainen, PhD Student, University of Jyväskylä, Finland
email:laura.k.kanniainen@jyu.fi

Professor Asko Tolvanen, University of Jyväskylä, Finland
email: asko.j.tolvanen@jyu.fi

Kaisa Lohvansuu, PhD, University of Jyväskylä, Finland
email: kaisa.lohvansuu@jyu.fi

Professor Paavo H.T. Leppänen, University of Jyväskylä, Finland
email: paavo.ht.leppanen@jyu.fi

Abstract
Identifying the factor structure of online reading to learn is important for the development of theory, assessment, and instruction. Traditional comprehension models have been developed from, and for, offline reading. This study used online reading to determine an optimal factor structure for modeling online research and comprehension among 426 sixth graders (ages 12 and 13). Confirmatory factor analysis was employed to evaluate an assessment of online research and comprehension based on a widely referenced theoretical model. Student performance reflected the theoretical constructs of the model, but several additional constructs appeared, resulting in a six-factor model: (a) locating information with a search engine, (b) questioning credibility of information, (c) confirming credibility of information, (d) identifying main ideas from a single online resource, (e) synthesizing information across multiple online resources, and (f) communicating a justified, source-based position. The findings are discussed in terms of theory, assessment, and instruction.

Theories of reading comprehension have long served to inform research, practice, and assessment in reading (e.g., Gates, 1921; Kintsch, 1998; Snow, 2002). All, however, have been developed from offline reading and for offline reading. As a result, they are both limited and limiting: They are limited to offline contexts, and they limit our ability to think in new ways about new forms of reading possible today. Today, reading takes place online in contexts that go beyond the reading of offline texts (Alvermann, 2002; Drew, 2012; Drounin & Davis, 2009) and in ways that may overlap but are not isomorphic with offline reading (Coiro, 2011). There is little theoretical and assessment work available to advance our understanding of reading to learn in these new contexts. New theories and new assessments are needed to capture the nature of reading online, especially reading to learn from online information as this ability is now important in every disciplinary area.

Anders, Yaden, Iddings, Katz, and Rogers (2016) called for greater attention to theory in the work we do as literacy researchers. Kintsch and Kintsch (2005) have suggested that an assessment conceptualized within a theory can inform both assessment and also our understanding of that theory. The purpose of this study is to explore theory by defining the factor structure of a new assessment of reading to learn from online information. The assessment was based on a theory of online research and comprehension developed from, and for, online contexts (Leu, Kinzer, Coiro, Castek, & Henry, 2013).

In this article, we first summarize research on reading to learn from online information. Next, we describe how this research is appearing in national standards and in international assessments. Then we describe a recent theory of online research and comprehension that was used to frame the assessment for this study. Finally, we present the factor structure that emerged from student performance and discuss the implications for theory, assessment, and instruction.

Reading to Learn From Online Information

The Internet has become a defining technology for reading and learning in these new times (Leu et al., 2013; Mills, 2010). As a result, researchers and policy makers have realized the need for understanding, assessing, and teaching the skills required to read and learn on the Internet (see Cho, Woodward, Li, & Barlow, 2017; Goldman, Braasch, Wiley, Graesser, & Brodowinska, 2012; Organisation for Economic Co-operation and Development [OECD], 2011). This work has become particularly important as students seem to be rather unskilled with reading to learn from online information (Bennett, Maton, & Kervin, 2008; Forzani, 2016; Kuiper, Volman, & Terwel, 2008).

Some forms of online reading appear to be different from offline reading, though the precise nature of the difference is still emerging. Afflerbach and Cho (2010) reviewed 46 studies focused on reading strategy use during Internet and hypertext reading. Their analysis showed evidence of strategies that "appear to have no counterpart in traditional reading" (p. 217). Many strategies centered around a reader's ability to apply methods to reduce his or her levels of uncertainty while navigating and negotiating appropriate reading paths in a shifting problem space (see also Zhang & Duke, 2008).

At least two lines of research have emerged in this area. First, research on reading to learn from online information has expanded our understanding of reading comprehension (e.g., Cho & Afflerbach, 2017; Goldman et al., 2012; Kiili, Laurinen, Marttunen, & Leu,

2012). Second, research on information problem solving online has expanded our

understanding of problem solving in complex environments (e.g., Brand-Gruwel, Wopereis,

& Vermetten, 2005; Brand-Gruwel, Wopereis, & Walraven, 2009).

Although these two lines of research have theorized somewhat similar component

structures, each uses slightly different component labels: (a) identifying the question/defining

the problem, (b) locating / searching for information, (c) evaluating/scanning information, (d)

synthesizing / processing information, and (e) communicating / presenting information. Many

of these components have appeared in think-aloud studies with small sample sizes (e.g.,

Brand-Gruwel et al., 2005; Coiro & Dobler, 2007). However, the complete factor structure

has never been explored with larger scale studies to establish the validity of these

components and structure.

This study was designed to examine the specific components derived from reading to

learn from online information to focus the evaluation on a well-specified theoretical model

(Leu et al., 2013).

Reading to Learn From Online Information in National Standards

The component structure studied in this investigation has found its way into

educational standards and curricula around the world, despite the fact that it has yet to be

tested. It is found, for example, in these boldfaced terms within an important design principle

of the Common Core State Standards in the United States (National Governors Association

Center for Best Practices & Council of Chief State School Officers, 2010).

To be ready for college, workforce training, and life in a technological society, students

need the ability to **gather**, comprehend, **evaluate**, **synthesize**, and **report** on

information and ideas, to **conduct original research in order to answer questions or**

solve problems, and to analyze and create a high volume and extensive range of print

and nonprint texts in media forms old and new. The need to conduct research and to

produce and consume media is embedded into every aspect of today's curriculum. (p. 4)

Australia, too, has included somewhat similar components in the Australia

Curriculum (Australian Curriculum Assessment and Reporting Authority, n.d.):

ICT competence is an important component of the English curriculum. . . . Students also

progressively develop skills in using information technology when **conducting**

research, a range of digital technologies to **create**, **publish and present their learning**,

and communication technologies to collaborate and **communicate with others** both

within and beyond the classroom. (para. 2)

In Finland, the context for this study, similar components also appear, especially in

two of the seven competency areas: information and communication technologies (ICT) and

multiliteracies (Finnish National Board of Education, 2014). For instance, the ICT section

identifies the importance of using search engines to find information from multiple resources,

critically evaluating information, and using multiple information resources in knowledge

construction, or synthesis. The multiliteracies section also emphasizes critical evaluation by

emphasizing the importance of understanding different purposes and perspectives of texts.

Many nations now include these components for reading to learn with online information

in their standards. This suggests the importance of evaluating a theoretical model and an

assessment framed around these components.

International Assessments of Reading to Learn From Online Information

Recently, two international assessments have provided useful direction for thinking

about the assessment of reading to learn from online information among adolescents: (a) the

Digital Reading Assessment of the Program for International Student Assessment, or PISA

(OECD, 2011), and (b) International Association for the Evaluation of Educational

Achievement's (IEA) International Computer and Information Literacy Study, or ICILS

(Fraillon, Ainley, Schulz, Friedman, & Gebhardt, 2014).

The Digital Reading Assessment (OECD, 2011) was designed to evaluate students'

skills in three areas: access and retrieve, integrate and interpret, and reflect and evaluate. The

ICILS (Fraillon et al., 2014) has two main dimensions. The first dimension focuses on

collecting and managing information that includes three aspects: knowing about and

understanding computer use, accessing and evaluating information, and managing

information. The second dimension includes producing and exchanging information.

Neither international assessment, however, evaluated students' ability to search for

information with a dynamic search engine that simulates actual keyword entry and the

generation of a results page. As work in information seeking online demonstrates (e.g., Bilal

& Kirby, 2002; Schacter, Chung, & Dorr, 1998), this is an essential aspect of reading to learn

from online information. The present study used statistical modeling for evaluating a

component structure derived from a theory that shared many of the component skills found in

these international assessments but included search engine use.

Theoretical Perspective

A theory of online research and comprehension (Leu, Kinzer, Coiro, & Cammack,

2004; Leu et al., 2013) was used to frame this study. Online research and comprehension is

defined as a self-directed process of constructing text and knowledge when seeking answers

to questions on the Internet. It uses the term *research* to establish the inquiry nature of most

reading on the Internet. It uses the term *comprehension* to establish its connection to

traditional comprehension models. This connection recognizes that reading on the Internet

includes a complex layering of offline and online comprehension processes. Online research

and comprehension comprises five components that are next briefly described.

Online research and comprehension is usually prompted by a question (Brand-Gruwel et al., 2005) that often directs the process (Zhang & Duke, 2008). In this study, we designed a prespecified activity that required students to work with a common question to standardize the task and permit reliable analyses.

Online readers typically locate relevant information with the help of search engines (Brand-Gruwel et al., 2009; Cho & Afflerbach, 2015). In the present study, a search engine was included in the assessment to evaluate students' ability with constructing an appropriate query (Guinee, Eagleton, & Hall, 2003) and analyzing search engine results (Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011).

The Internet, with information that varies widely in quality, challenges readers to critically evaluate the credibility of online information (Fabos, 2008; Forzani, 2016; Wineburg, McGrew, Breakstone, & Ortega, 2016). When evaluating credibility, skilled readers pay attention to at least two aspects of the resource: expertise and trustworthiness (Flanagin & Metzger, 2008; Tseng & Fogg, 1999). To evaluate varying types of online information, this study asked students to evaluate two types of resources: an academic, somewhat neutral, resource and a commercial, somewhat biased, resource.

The ability to synthesize, or integrate, ideas from multiple online resources is a particularly important aspect of online research and comprehension (DeSchryver, 2015; Goldman et al., 2012). Readers select important ideas from multiple resources, and they organize, connect, compare, and contrast these ideas to build a coherent understanding of the issue (Cho & Afflerbach, 2015; Rouet, 2006; Strømsø, Bråten, & Britt, 2010). In the present study, we evaluated synthesis by asking students to first take notes from four online resources that provided different perspectives and, in some cases, conflicting information about the issue and then integrate information from these notes.

Online research and comprehension conceives of a close connection between the reading and the writing that takes place during online communication. Studies investigating online inquiry processes often ask students to communicate their understanding with an essay (Goldman et al., 2012; Le Bigot & Rouet, 2007). In the current study, students were asked to communicate a recommendation, with justifications, to a fictitious school principal in an email. This allowed us to evaluate students' abilities to take and justify a position and also their sensitivity to audience.

Research Question

This study sought to determine the factor structure for the online research and comprehension process. Specifically, we used the following research question to direct the investigation:

What is the optimal factor structure for a performance-based assessment that was conceptualized within a theory of online research and comprehension?

The answer can provide useful information about both the validity of the assessment and the theory on which it is based (Kintsch & Kintsch, 2005), informing us about the important classroom issue of reading to learn from online information.

Method

Participants

Participants in this study were 426 sixth graders (219 boys and 207 girls) ages 12 and 13. Students were recruited from 24 classes representing eight Finnish elementary schools that included both suburban and rural schools. All students spoke Finnish as their first language. Almost all students had access to the Internet at home: 99% of students reported having home access via desktop computer, laptop, or tablet, and 96% reported having home access via smartphone. While family income data are not available, participating students represented the typical range of socioeconomic status in Finland in terms of highest parental

education: 40% of students had at least one parent with university education, 26% had at least

one parent with a university of applied sciences education, 33% had both parents with

vocational education or a high school education, and 2% of students' parents did not have

further education after comprehensive schooling. Generally, students reflected the average

range of reading ability found in Finland in both fluency and reading comprehension.

Online Research and Comprehension Measure

Students' skills were measured with an "**I**nternet**l**ukemisen **a**rviointitesti (ILA),"

which is the Finnish translation for online research and comprehension assessment. This was

modified from a previous assessment that was developed with acceptable levels of reliability

and validity during a 5-year, federally funded research grant in the United States (see

Kennedy, Rhoads, & Leu, 2016; Leu, Coiro, Kulikowich, & Cui, 2012; Leu et al., 2015). The

Kuder–Richardson Formula 20 (KR-20) estimates of reliability for this assessment ranged

from .70 to .82 (see Kennedy et al., 2016; Leu et al., 2015). Validity was established with a

framework document approved by experts, 2 years of cognitive lab testing, and modifications

based a large-scale pilot study. While target online resources for each search problem were

specified in the programming algorithm in advance, unlike actual reading online, this

permitted a standardized task for evaluating performance. The ILA was adapted from this

prior assessment and translated into Finnish. Some modifications in word choice were made

in the instructions and scaffolding prompts to better reflect the Finnish language and culture.

Both assessments measured the following components of online research and comprehension,

consistent with the theoretical model: locating information, evaluating information,

synthesizing information, and communicating information. Neither assessment included a

task to measure the component of identifying important questions; all students were asked to

work with the same prespecified question. Working to answer the same question ensured that

we were able to compare students' performance on a common task.

Students were given an online research task in a virtual online space where they were guided by an avatar. They received an email message from a school principal, which asked them to use online information to explore the health effects of energy drinks and provide a recommendation as to whether or not she should allow an energy drink machine to be placed in the school. To compose an informed recommendation, students read four online resources that were within their reading ability. The texts explored the health effects of energy drinks from different perspectives. Two online resources were given to students and students were asked to locate two others with a search engine. In addition, students were later asked to evaluate two of these resources in relation to the authors' expertise and the overall credibility of resources. The flow of the task is presented in Figure 1. Furthermore, Table 1 summarizes the characteristics of each online resource and indicates whether students were initially given a resource or asked to locate it. It also shows which resources students were asked to evaluate.

--FIGURE 1--

--TABLE 1--

Two changes were made to the ILA. First, instead of evaluating the credibility of a single online resource, students evaluated two resources that varied in purpose, argumentation, and author's expertise. Second, instead of three partial summaries from separate resources, students composed a final summary of their knowledge from all resources. Students in Finland had a limited amount of time (1 class hr) to complete the test, and we wanted to provide sufficient time to compose the final summary.

Data Collection

Data were collected at schools during a 45-min period during regular classes. Students worked with identical laptops, and the assessment was run from the local server of each laptop. As we had a limited number of laptops, half of the class worked on the task at a time.

This ensured that students could be placed in the classroom so that they were not able to see other students' screens. Students proceeded at their own pace and, if needed, they were allowed to use their 15-min recess to complete the task. If students encountered technical problems, the researchers helped them to proceed with the task, but only solved technical issues. These were infrequent.

Scoring Students' Performance in the Online Research Task

Table 2 provides the scoring rubric used to measure each skill. Score ranges varied in accordance with the complexity of the students' answers or activities. For example, the shorter range (0-2 points) was used to score students' keyword queries, whereas the longer range (0-5 points) was used to score variables related to communicating via email, where the longest email was as long as 183 words. Table 2 also includes estimates of interrater reliability. The first coder scored all the responses, after which the second coder scored 20% of the responses from randomly selected participants for all variables except for the measurement of time spent on locating the correct page, which was auto scored. Kappa values ranged from 0.722 to 1.000. Table 3 includes a detailed description of how each item was scored.

--TABLE 2--

--TABLE 3--

In this study, the system provided students with the correct online resource if they were unable to locate it in 3 minutes. In spite of this feature, some students did not realize (or show concern) that the resource they selected and read was not the correct one. This resulted in the possibility that these students would not receive scores from the notes that they were asked to take, related to the target resources that they failed to locate (Resources 2 and 4). To minimize the error, students who did not take notes from the correct resources were not included in the analysis of the variable NOTE2 ($n = 41$, or 9.2%) or NOTE4 ($n = 24$, or

5.6%) or both of them ($n = 9$, or 2.1%). The inability to locate a correct resource may have

had some influence on scores in the use of sources (SUM1) and the quality of the summary

(SUM2). In addition, location issues may have influenced the quality of argumentation in

email messages (COM1) because these students may have been exposed to less of the

relevant content compared with students who had an opportunity to read all four of the target

online resources. These consequences are somewhat similar to what often happens during

online research and comprehension, when performance at early stages often affects later

stages of performance (see Henry, 2006).

Statistical Analyses

Confirmatory factor analysis (CFA) was employed to assess the measurement model

of online research and comprehension. Mplus 7.4 software was used. A variance-adjusted

weighted least squares (WLSMV) estimation method was employed because categorical

variables were included in the model and six of them were not normally distributed. Also, the

WLSMV estimator has been demonstrated to provide less biased estimates for factor loadings

with categorical variables (Li, 2016), which further supported its use.

The starting point for CFA modeling was the four-factor model of online research

and comprehension (Leu et al., 2013). It was hypothesized that alternative models could exist

for online research and comprehension measured with the ILA. The following fit indices and

cutoff values were used for indicating goodness of fit: χ^2-test (ns, $p > .05$), root mean square

error of approximation (RMSEA) values $\leq .06$, and Tucker–Lewis index (TLI) and

comparative fit index (CFI) values $\geq .95$ (Hu & Bentler, 1999).

Should the four-factor model show lack of fit, it was planned to continue the CFA by

adopting a model generation approach. In this approach, the aim is to establish a model that is

acceptable in terms of both statistics and theory (Kline, 2011). Here, the hypothesized

measurement model of online research and comprehension originated from the current

theory. Therefore, the approach suited our purpose, which was to show the possible

incongruity between our data and the hypothesized model. This process was iterative, where

the model was altered in each step based on modification indices and on the theoretical

justification until a sufficient model was established. In the last step, the resulting, less

restrictive, model was tested against the more restrictive one using the chi-square difference

($\chi^{2\text{-diff}}$-test) test. When comparing these two nested models, the DIFFTEST-option,

implemented in Mplus, was employed.

Findings

To estimate the best measurement model for this type of assessment, a simulation

within an online space, we initially estimated a four-factor model for online research and

comprehension, based on existing theory. All parameter estimates (see Figure 2) were

statistically significant (ps < .01). Similarly, all fit indices except the χ^2-test indicated good

model fit, $\chi^2(84) = 172.20$, $p < .001$; RMSEA = .05; CFI = .97; TLI = .96. However,

modification indices suggested that allowing the residuals of SUM1 and SUM2 to correlate

would improve the model. Instead of estimating the residual correlation, the *Synthesizing*

factor was divided into two factors: *Identifying Main Ideas* and *Synthesizing*.

--FIGURE 2--

Data in this study suggest that the act of composing a summary is based on the ability

to integrate multiple ideas collected when taking notes, but it appears to consist of two,

somewhat separate, elements. First, composing a summary appears to require one to identify

and take notes on the main ideas from each online resource. Second, it requires one to

integrate or synthesize the main ideas from multiple resources into a cohesive summary.

These findings are consistent with research that differentiates between the processes of

determining important ideas and synthesizing important ideas within and across static printed

texts (Pressley & Afflerbach, 1995) and more dynamic digital texts (Cho & Afflerbach, 2017).

In line with the results of four-factor model analysis, the five-factor model (see Figure 3), with the factors of *Identifying Main Ideas* and *Synthesizing*, demonstrated good fit to the data. Fit indices were above or under the cutoff values, except for the χ^2-test, which was significant, $\chi^2(80) = 143.15$, $p < .001$; RMSEA = .04; CFI = .98; TLI = .97. The χ^2-difference test indicated that the less restricted five-factor model would fit the data better, $\chi^{2\text{-diff}}(4) = 23.60$, $p < .001$, than the four-factor model. Yet, the modification indices suggested that the model fit would be better if the residuals of variables EV1 and EV2 were allowed to correlate. Thus, instead of estimating residual correlation, the *Evaluating* factor was divided into two factors: EV1 + EV2 and EV3 + EV4. From a theoretical perspective, this means that evaluating an academic, somewhat neutral, resource (measured by Items EV1 and EV2) and evaluating a commercial, somewhat biased, resource (measured by EV3 and EV4) may require somewhat different abilities.

<div align="center">--FIGURE 3--</div>

As pointed out earlier, the academic, somewhat neutral, resource in our study included information that did not take a strong stance, whereas the commercial, somewhat biased, resource contained information that was one-sided. Consequently, Items EV1 and EV2 mainly required students to confirm the credibility of a resource, while Items EV3 and EV4 required students to question a resource's credibility by identifying, for example, its commercially driven one-sided stance or by identifying a conflict with information from other resources. These results are consistent with work that indicates that evaluating the credibility of resources is often more complex than may initially appear and may be associated with both confirming and questioning resources (Klayman, 1995; Masnick & Zimmerman, 2009). As the existence of separate factors (*Confirming Credibility* and

Questioning Credibility) seemed to be both statistically and theoretically acceptable, the six-

factor measurement model of online research and comprehension was tested.

Results (see Figure 4) showed that the six-factor model fit the data very well. All

goodness-of-fit indices, including the χ^2 test, fulfilled the criteria for a good model fit.

Moreover, the χ^2-difference test showed that the six-factor model had a significantly better fit

than the five-factor model, $\chi^{2\text{-diff}}(5) = 43.08$, $p < .001$. Based on these results, the six-factor

model of online research and comprehension was considered as the final model that best fit

these data.

<div align="center">--FIGURE 4 –</div>

A comparison of the three measurement models reveals that, by specifying the new

factors to the model, additional interpretive power becomes available without substantial loss

and, in some cases, with an increase in the relationships between the factors. The

interpretations appear to be consistent with existing theory and research. Factor correlations,

for example, appeared to differ in an interpretable direction for at least one of the new factors

from *Synthesizing*. When the *Identifying Main Ideas* factor was specified in the five- and six-

factor models, it had a higher correlation ($r = .73$) with the *Synthesizing* factor and a lower

correlation ($r = .58$) with the *Communicating* factor. This makes some sense, suggesting that

being able to identify the main ideas of an online resource contributes somewhat more to the

prediction of synthesis, a more focused task, than the prediction of communication, which

required students to also use email tools and appropriate discourse to compose their

arguments for a particular audience. This gain in explanatory power did not appear to come at

the expense of the overall relationship between *Synthesizing* and *Communicating*. In the four-

factor model, the correlation between the *Synthesizing* and *Communicating* factors was .67.

The correlation between *Synthesizing* and *Communicating* was .66 in both the five- and six-

factor models, even though the *Identifying Main Ideas* factor appeared separately in the five-
and six-factor models.

Separating critical evaluation into two factors, *Confirming Credibility* and
Questioning Credibility, resulted in stability between relationships when compared with the
four- and five-factor models. In the four- and five-factor models, correlations between
Evaluating and *Locating* factors were .33. In the six-factor model, the correlations of
Confirming Credibility and *Questioning Credibility* factors with the *Locating* factor were
both .29. Thus, relationships with *Locating* did not change substantially when the two
separate factors for *Evaluating* were specified in the six-factor model.

A similar pattern appeared for correlations with *Communicating*. In the four-factor
model, correlations for *Evaluation* with *Communicating* were .57. In the six-factor model, the
correlations of *Confirming Credibility* and *Questioning Credibility* with *Communicating* were
.52 and .48, respectively. These associations also did not change substantially with the gain in
explanatory power by adding additional components into the six-factor model.

The pattern was slightly different in relation to *Evaluating* and *Synthesizing* when
both factors were expanded in the six-factor model. In the four-factor model, the association
between *Evaluating* and *Synthesizing* was .52. In the six-factor model, correlations between
Confirming Credibility with *Identifying Main Ideas* and *Synthesizing* did not change
substantially, as they were .57 and .47, respectively. Correlations decreased substantially,
however, between *Questioning Credibility* with *Identifying Main Ideas* and *Synthesizing*.
These were .46 and .33, respectively. Thus, separating both *Evaluating* and *Synthesizing* into
additional factors generally sustained the relationships for *Confirming Credibility*. For
Questioning Credibility, however, the additional explanatory power, by adding factors, came
at the cost of somewhat lower associations with the factors of *Synthesizing*.

Looking at all of the consequences, resulting from the additional factors in the six-factor model, results indicate that estimating two *Evaluating* factors and estimating separate *Identifying Main Ideas* and *Synthesizing* factors add important information to our understanding of online research and comprehension while generally sustaining the overall structure of the theoretical model.

Discussion

This study sought an optimal component structure for a theoretical model of online research and comprehension. The results are discussed in terms of contributions to theory, assessment, and instruction.

Theory

Factor analytic studies have the potential to clarify and refine theoretical models. Results show that students' performance measured with the ILA assessment generally reflected the original theoretical constructs of online research and comprehension. The four measured component skills of locating, evaluating, synthesizing, and communicating information appeared as separate constructs with a good model fit. This suggests some support for the overall structure of the theoretical model. However, the five- and six-factor models showed even better fit by including two factors for each of the evaluation and synthesis components. This added additional complexity to the model by including six, not four, factors: (a) locating information with a search engine, (b) confirming the credibility of information, (c) questioning the credibility of information, (d) identifying main ideas from a single online resource, (e) synthesizing information across multiple online resources, and (f) communicating a justified and source-based position.

It seems that the evaluation of different types of resources may require somewhat different types of evaluation skills as the single factor, evaluation, separated into two factors that aligned with the two different contexts used to measure evaluation. An academic

resource presenting facts in a neutral manner appeared to require confirmation of its overall

credibility and thus its suitableness as a learning resource. A commercial resource with one-

sided argumentation appeared to require questioning of its overall credibility. It is not clear

from the results of this study if a wider range of resources, possibly requiring additional

forms of evaluation, would separate into additional factors or coalesce back around a single

factor. The moderately high correlation ($r = .49$, $p < .001$) between confirming the credibility

of information and questioning the credibility of information suggests that there is some

overlap in these skills. Two factors reflecting the different types of critical evaluation skills

might be specific to younger students compared with expert readers (cf. Brand-Gruwel et al.,

2005). Further studies are needed to explore whether these results hold in different disciplines

(Goldman et al., 2016) and with a broader range of resources, and with older, more

cognitively mature students (cf. Eccles, Wigfield, & Byrnes, 2003).

  The synthesis component also separated into two factors, aligning with each synthesis

task: identifying main ideas from a single online resource and synthesizing information

across multiple resources. This is consistent with what Cho and Afflerbach (2017) have

described as the dual processes involved in building multilevel coherence while reading

digital resources. They argue that the process of building coherence within a single online

text (in this case, taking notes from each online resource) is separate from that of building

coherent intertextual relationships across multiple online texts (in this case, completing the

final synthesis task).

  These results also suggest that the dual processes may be interdependent. Identifying

main ideas from a single resource and synthesizing information across resources were highly

correlated ($r = .73$, $p < .001$). This interdependence may have at least two explanations. First,

similar types of reading strategies may be needed when constructing meaning from a single

resource and across multiple resources. For example, a reader may strategically generate

inferences about connections while reading within a single resource as well as when reading across multiple resources. Second, the identification of main ideas from a single resource forms the basis for the synthesis of ideas across multiple resources (Cho & Afflerbach, 2017).

The high factor loadings on items designed to measure the ability to communicate on the basis of one's research suggests these two items effectively captured the component skill of communicating via email as part of online research and comprehension. Notably, the communicating factor correlated with identifying main ideas from a single online resource (r = .58, p < .001) and synthesizing information across multiple resources (r = .66, p < .001). These associations suggest that success in previous phases of the task may have assisted students with communicating a position in their email, supporting conceptions about intrinsic relationships between reading and writing during online research and comprehension.

Taken together, the results appear to not only provide some support for the initial model but also suggest that additional complexities may be an essential aspect of any final model. The use of CFA by adopting the model-generating approach and comparing the models via the chi-square difference test was found to be feasible. This approach is useful for adjusting theoretically well-justified models and then comparing them against each other. In future studies, measurement invariance testing of the online research and comprehension model would be highly recommended as it reveals whether the constructs of the model were measured in the same manner within the different groups of students. This study found at least two of the areas to divide into separate factors when two different types of tasks were required for each. We note, however, the need to be cautious as the topic (food and health) is one subject in school learning at this level but it is only one of countless areas of learning. Also, as new technologies appear online and new task demands appear, adding complexity to an already complex context for reading, more complexities in models of online research and comprehension may be revealed.

None of the factors identified here, with the exception of identifying main ideas from a single online resource and possibly synthesizing information across multiple online resources, typically appear in traditional offline reading models. From a metatheoretical perspective, these results lead us to reconsider traditional models of reading developed from and for offline reading contexts. The results of this study suggest that theoretical models grounded in offline reading contexts are likely to ignore elements distinctive to online reading as well as the differential functioning of common elements during reading in online and offline contexts (Leu et al., 2015). This could lead us in inappropriate directions as we consider the design of curriculum, instruction, and standards for online reading.

The model that emerged in this study, developed from and for online reading contexts, includes a number of components that have not previously been identified as central to offline models, suggesting transformative possibilities as we envision reading in new ways for new times. Theoretically, it appears that we must begin to recognize both the distinctive nature as well as the central importance of the online informational contexts that now define our literacy lives (cf. Finnish National Board of Education, 2014; Fraillon et al., 2014; International Reading Association, 2009). Our findings suggest that it is problematic to simply adapt offline models to online contexts. It is better to develop theory from the reality that emerges from online contexts.

Results that diverge from offline models may also provide important direction for online theory development in additional areas such as multiple document comprehension (e.g., Bråten & Braasch, 2017; Britt & Rouet, 2012), multimodality (e.g., Smith, Kiili, & Kauppinen, 2016), and critical literacy (e.g., Lankshear & McLaren, 1993), and others. This study helps us to consider the problematic aspects of simply adapting offline theory to new, online contexts in these areas and the importance of developing new theoretical perspectives from within the many new, online literacy contexts themselves (see Leu et al., 2013).

Finally, these theoretical results may provide additional support for repositioning online reading to a central location in the language arts (see Kervin, Mantei, & Leu, 2018). It is no longer sufficient to continue to position online reading instruction as secondary, supplemental, and tangential to offline reading instruction, seemingly as an afterthought. Using online information to read, solve problems, and learn is now such an integral part of our lives (OECD, 2011) that it must become the central focus of the thoughtful preparation of the next generation. The model and components identified in this study may provide important initial direction to the changes required in classrooms. Such a repositioning of online reading instruction to a central position in our classrooms will severely test our ability to prepare students for literacy, learning, and life in this online age of information and communication, and we need a clear understanding of online reading to be successful.

Assessment

Results demonstrating a factor structure for an assessment generally similar to the theory behind it suggests a reasonable level of validity for that assessment (Rahman & Mislevy, 2017). This seems true for this instrument. Adding to its validity is that this was a performance-based assessment, seeking to replicate, as closely as possible, how individuals read to learn online, conduct research, and comprehend issues related to health. This approach may have advantages over assessments that use isolated, multiple-choice items (cf. Sabatini, O'Reilly, Halderman, & Bruce, 2014) as it allows the capture of a more integrated set of competencies associated with higher level thinking when reading to learn from online information. A simulation may have also increased students' engagement in the task (de Klerk, Veldkamp, & Eggen, 2015).

With respect to measuring locating skills, the ILA assessment included a search engine designed to simulate a dynamic search environment in which students were able to formulate and revise queries that directly informed what was displayed on a subsequent

search result page. This made it possible to compare students' locating performance because they could all engage with the same set of online resources. Assessment designers seeking to authentically capture dynamic locating skills, while also holding constant some elements of the information space, might consider integrating similar features into future assessments.

Despite these benefits, our approach to designing the four locating items also introduced some complexities. First, we found that the two query-formulation tasks inconsistently measured students' locating skills. Consequently, we decided to exclude one of the query-formulation items from the factor analysis. This was because it appeared that the query-formulation tasks required students to use different types of querying strategies; that is, a "topic + defining concept query" was useful for locating information about a specific topic (LOC2) but a "topic + source query" was more appropriate for locating information published by a certain author/organization (LOC1). While this created challenges for fitting the four items to a common locating factor, the varied difficulty levels of these two items, LOC1 being more difficult for students, also suggests that there might be some nuanced levels of locating skills related to query formulation in different contexts. Future assessments might include more items that necessitate both kinds of queries to more carefully explore this possibility.

Additional research is also needed to clarify whether or not locating tasks that involve query formulation and locating tasks that involve the selection of relevant links from the search results page measure different underlying abilities. Previous research (Cho & Afflerbach, 2015; Coiro & Dobler, 2007) found that searching tasks require prior knowledge of Web-based search engine protocols (i.e., formulating keyword searches, evaluating annotated search results) as well as inferential reasoning skills (i.e., making, confirming, and adjusting inferences in relation to context cues). However, less is known about the extent to

which query-generating and link-selection skills overlap or separate into different factors involving slightly different sets of processes.

Finding a balance between ecological validity (authenticity) and standardization is a key issue in the design of performance-based assessments (Linn, Baker, & Dunbar, 1991). Reading in the authentic Internet environment involves a significant degree of student choice. There is, however, a need to limit these choices somewhat to compare students' performance on a common, or standardized, task. This requirement resulted in two important limitations to the study.

First, as students' performance was standardized by providing an identical question to each student, we were not able to measure one component of the theoretical model for online research and comprehension: identifying important questions. As online research and comprehension is a self-directed reading process, identifying important questions appears to not just influence the beginning of an online research and comprehension task but may also influence the process at various points as the reader revises, restructures, and reconstructs the question on the basis of new information that he or she acquires (see Cho et al., 2017). Future research needs to find creative solutions to embed this component into assessments without jeopardizing the comparability of students' performance. Second, while the assessment was designed only to simulate the Internet, standardizing the task required us to hold some of the elements in the information space constant (e.g., resources that students eventually read, the sequence of the tasks, and prompts from the avatar).

All these design decisions compromised a part of the authenticity of online research and comprehension as a self-directed reading process where each reader creates a unique reading path in an unlimited information space. Consequently, this study is limited by having had to standardize the assessment task in these ways. We note, however, that compromises to authenticity also appear in standardized assessments of offline reading where students do not

select their own reading material and typically read standard portions of limited texts while responding to a limited and fixed set of possible responses. More flexible assessment environments and more creative solutions to the ones used in this assessment are needed to more closely match authentic online research and comprehension.

In spite of these limitations, this study gives important direction for developing assessments of reading to learn on the Internet. In addition to including a search engine designed to simulate a dynamic search environment, another important aspect that has not yet been fully included in the international assessments is the synthesis of information from multiple online texts. Research on multiple document literacy (Bråten & Braasch, 2017; Britt & Rouet, 2012) has highlighted the importance of reading multiple texts representing different, even conflicting perspectives and voices, as well as the unique skills that are needed beyond the reading of a single text. Future assessments should also evaluate students' abilities to handle information that is biased, prejudiced, or intentionally misleading, now needed in what some might call our "post-factual era." As the present study showed, this might require additional skills in critical literacy with online information at least from younger students.

Instruction

The use of performance-based assessments, systematically aligned with instruction, can provide important support for teachers (Pellegrino, DiBello, & Goldman, 2016). This may be especially true with this assessment as the component skills appear in many nations' standards (see Australian Curriculum Assessment and Reporting Authority, n.d.; Finnish National Board of Education, 2014; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). This study found six specific components related to online research and comprehension that appear important for success

in reading to learn from online information. Understanding the importance of these

components may lead directly to the design of more effective instructional activities.

  The six-factor model of online research and comprehension found in this study can

also help clarify the similarities and differences between the skills associated with each

factor. For example, knowing that questioning an author's credibility requires different skills

than confirming an author's credibility suggests the need to more strategically sequence

instruction around critically evaluating online resources. Students need both models and

practice in considering what may cause someone to question an author's credibility. To avoid

superficial practices, such as making evaluative judgments only on the basis of domains

(.edu, .org), students should be pointed back to the information itself to analyze the quality of

an author's claims about the topic under study in relation to his or her affiliation and level of

expertise. Having regular opportunities to engage with challenging critical evaluation

practices is likely to increase students' ability to identify possible persuasive elements and

one-sided arguments that call into question the credibility of many online texts (cf. Buehl,

Alexander, Murphy, & Sperl, 2001).

  Similarly, this study highlighted two separate but related skills that can inform

instruction in effectively synthesizing ideas across multiple online texts. Like previous

research (Cho & Afflerbach, 2015; Rouet, 2006; Spivey & King, 1989), our findings suggest

that online readers need to first be able to determine the most important ideas within a single

resource relevant to the task at hand. Then, they need to integrate these separate ideas from

several resources to form cohesive explanations. It appears likely that working to integrate

information across multiple and disparate resources may introduce new challenges for

students more accustomed to reading logically structured learning materials, such as a

textbook or a single article. Thus, instructional supports need to go beyond making sense of

single texts to help students understand how multiple resources, especially from different

perspectives, contribute to the overall understanding of a topic (Fisher & Frey, 2013; Hynd, Holschuh, & Hubbard, 2004).

Acknowledgments

Funding

References

Afflerbach, P., & Cho, B.-Y. (2010). Determining and describing reading strategies: Internet and traditional forms of reading. In H. S. Waters & W Schneider (Eds.), *Metacognition, strategy use, and instruction* (pp. 201-225). New York, NY: Guilford Press.

Alvermann, D. (2002). Effective literacy instruction for adolescents. *Journal of Literacy Research, 34*, 189-208. doi:10.1207/s15548430jlr3402_4

Anders, P. L., Yaden, D. B., Iddings, A. C., Katz, L., & Rogers, T. (2016). Improving theory in literacy research. *Journal of Literacy Research, 48*, 3-4. doi:10.1177/1086296X16637169

Australian Curriculum Assessment and Reporting Authority. (n.d.). *The Australian Curriculum, v6.0*. Retrieved from www.australiancurriculum.edu.au/Home

Bennett, S., Maton, K., & Kervin, L. (2008). The "digital natives" debate: A critical review of the evidence. *British Journal of Educational Technology, 39*, 775-786. doi:10.1111/j.1467-8535.2007.00793.x

Bilal, D., & Kirby, J. (2002). Differences and similarities in information seeking: Children and adults as Web users. *Information Processing & Management, 38*, 649-670.

doi:10.1016/

s0306-4573(01)00057-7

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by

experts and novices: Analysis of a complex cognitive skill. *Computers in Human

Behavior, 21*, 487-508. doi:10.1016/j.chb.2004.10.005

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information

problem solving while using Internet. *Computers & Education, 53*, 1207-1217.

doi:10.1016/j.compedu.2009.06.004

Bråten, I., & Braasch, J. L. G. (2017). Key issues in research on students' critical reading and

learning in the 21st century information society. In C. Ng & B. Bartlett (Eds.), *Improving

reading and reading engagement in the 21st century: International research and

innovations* (pp. 77-98). Singapore: Springer.

Britt, M. A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and

their acquisition. In M. J. Lawson & J. R. Kirby (Eds.), *Enhancing the quality of learning:

Dispositions, instruction, and learning processes* (pp. 276-314). New York, NY:

Cambridge University Press.

Buehl, M. M., Alexander, P. A., Murphy, P. K., & Sperl, C. T. (2001). Profiling persuasion:

The role of beliefs, knowledge, and interest in the processing of persuasive texts that vary

by argument structure. *Journal of Literacy Research, 33*, 269-301.

Cho, B.-Y., & Afflerbach, P. (2015). Reading on the Internet. *Journal of Adolescent & Adult

Literacy, 58*, 500-513. doi:10.1002/jaal.387

Cho, B.-Y., & Afflerbach, P. (2017). An evolving perspective of constructively responsive

reading comprehension strategies in multilayered digital text environments. In S. Israel

(Ed.), *Handbook of research on reading comprehension* (2nd ed., pp. 109-134). New York,

NY: Guilford Press.

Cho, B.-Y., Woodward, L., Li, D., & Barlow, W. (2017). Examining adolescents' strategic processing during online reading with a question-generating task. *American Educational Research Journal, 54*, 691-724. doi:10.3102/0002831217701694

Coiro, J. (2011). Predicting reading comprehension on the Internet: Contributions of offline reading skills, online reading skills, and prior knowledge. *Journal of Literacy Research, 43*, 352-392.

Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly, 42*, 214-257. doi:10.1598/RRQ.42.2.2

de Klerk, S., Veldkamp, B. P., & Eggen, T. J. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & Education, 85*, 23-34. doi:10.1016/j.compedu.2014.12.020

DeSchryver, M. (2015). Web-mediated knowledge synthesis for educators. *Journal of Adolescent & Adult Literacy, 58*, 384-392. doi:10.1002/jaal.373

Drew, S. V. (2012). Open up the ceiling on the common core state standards: Preparing students for 21st-century literacy—Now. *Journal of Adolescent & Adult Literacy, 56*, 321-330. doi:10.1002/JAAL.00145

Drounin, M., & Davis, C. (2009). R u texting? Is the use of text speak hurting your literacy? *Journal of Literacy Research, 41*, 46-67. doi:10.1080/10862960802695131

Eccles, J. S., Wigfield, A., & Byrnes, J. (2003). Cognitive development in adolescence. In R. Lerner, M. Easterbrooks, & J. Mistry (Eds.), *Handbook of psychology. Vol. 6. Developmental psychology* (pp. 325-350). New York, NY: Wiley.

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

Finnish National Board of Education. (2014). *Finnish national curriculum for basic*

education. Retrieved from http://www.oph.fi/english/curricula_and_qualifications/

basic_education

Fisher, D., & Frey, N. (2013). *Reading and reasoning: Fostering comprehension across*

multiple texts. Retrieved from http://www3.canisius.edu/~justice/CSTmodule-final/ira-e-

ssentials-8026-reading-and-reasoning.pdf

Flanagin, A. J., & Metzger, M. J. (2008). Digital media and youth: Unparalleled opportunity

and unprecedented responsibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media,*

youth, and credibility (pp. 5-27). Cambridge, MA: MIT Press.

Forzani, E. (2016). *Individual differences in evaluating the credibility of online information*

in science: Contributions of prior knowledge, gender, socioeconomic status, and offline

reading ability (Unpublished doctoral dissertation). University of Connecticut, Storrs.

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2014). *Preparing for life in*

a digital age: The IEA International Computer and Information Literacy Study

International Report. Melbourne, Australia: Australian Council for Educational Research

(ACER). Retrieved from

http://www.iea.nl/fileadmin/user_upload/Publications/Electronic_versions/

ICILS_2013_International_Report.pdf

Gates, A. I. (1921). An experimental and statistical study of reading and reading tests.

Journal of Educational Psychology, 12, 445-464. doi:10.1037/h0064253

Goldman, S. R., Braasch, J. L. G., Wiley, J., Graesser, A. C., & 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., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., & Shanahan,

C. (2016). Disciplinary literacies and learning to read for understanding: A conceptual

framework for disciplinary literacy. *Educational Psychologist, 51*, 219-246.

doi:10.1080/00461520.2016.1168741

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.

Henry, L. A. (2006). SEARCHing for an answer: The critical role of new literacies while

reading on the Internet. *The Reading Teacher, 59*, 614-627. doi:10.1598/rt.59.7.1

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure

analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling,

6*(1), 1-55. doi:10.1080/10705519909540118

Hynd, C., Holschuh, J. P., & Hubbard, B. P. (2004). Thinking like a historian: College

students' reading of multiple historical documents. *Journal of Literacy Research, 36*, 141-

176. doi:10.1207/s15548430jlr3602_2

International Reading Association. (2009). *IRA position statement on new literacies and 21st

century technologies*. Newark, DE: Author. Retrieved from

http://www.literacyworldwide.org/docs/

default-source/where-we-stand/new-literacies-21st-century-position-statement.pdf?sfvrsn=6

Kennedy, C., Rhoads, C., & Leu, D. J. (2016). Online research and learning in science: A

one-to-one laptop comparison in two states using performance based assessments.

Computers & Education, 100, 141-161. doi:10.1016/j.compedu.2016.05.003

Kervin, L., Mantei, J., & Leu, D. J. (2018). Repositioning online reading to a central location

in the language-arts. In D. Lapp & D. Fisher (Eds.), *Handbook of research on teaching the

English language arts* (4th ed., pp. 328-359). New York, NY: Routledge.

Kiili, C., Laurinen, L., Marttunen, M., & Leu, D. J. (2012). Working on understanding during

    collaborative online reading. *Journal of Literacy Research, 44*, 448-483.

    doi:10.1177/1086296X12457166

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge

    University Press.

Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. G. Paris & S. A. Stahl (Eds.),

    Children's reading comprehension and assessment (pp. 71-92). Mahwah, NJ: Lawrence

    Erlbaum.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation,

    32*, 385-418. doi:10.1016/S0079-7421(08)60315-1

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY:

    Guilford Press.

Kuiper, E., Volman, M., & Terwel, J. (2008). Students' use of Web literacy skills and

    strategies: Searching, reading and evaluating Web information. *Information Research,

    13*(3), Paper 351. Retrieved from http://InformationR.net/ir/13-3/paper351.html

Lankshear, C., & McLaren, P. (Eds.). (1993). *Critical literacy: Politics, praxis, and the

    postmodern*. Albany: State University of New York Press.

Le Bigot, L. L., & Rouet, J. F. (2007). The impact of presentation format, task assignment,

    and prior knowledge on students' comprehension of multiple online documents. *Journal of

    Literacy Research, 39*, 445-470. doi:10.1080/10862960701675317

Leu, D. J., Coiro, J., Kulikowich, J., & Cui, W. (2012, December). *Using the psychometric

    characteristics of multiple-choice, open Internet, and closed (simulated) Internet formats

    to refine the development of online research and comprehension assessments in science:

    Year three of the ORCA project*. Paper presented at the 62nd, annual meeting of the

    Literacy Research Association, San Diego, CA.

Leu, D. J., Forzani, E., Rhoads, C., Maykel, C., Kennedy, C., & Timbrell, N. (2015). The

new literacies of online research and comprehension: Rethinking the reading achievement

gap. *Reading Research Quarterly, 50*, 37-59. doi:10.1002/rrq.85

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

literacies emerging from the Internet and other information and communication

technologies. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of

reading* (5th ed. pp. 1568-1611). Newark, DE: International Reading Association.

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2013). New literacies and the

new literacies of online reading comprehension: A dual level theory. In N. Unrau & D.

Alvermann (Eds.), *Theoretical models and process of reading* (6th ed., pp. 1150-1181).

Newark, DE: International Reading Association.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum

likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*, 936-

949. doi:10.3758/s13428-015-0619-7

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment:

Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.

Masnick, A., & Zimmerman, C. (2009). Evaluating scientific research in the context of prior

belief: Hindsight bias or confirmation bias. *Journal of Psychology of Science and

Technology, 2*(1), 29-36. doi:10.1891/1939-7054.2.1.29

Mills, K. A. (2010). A review of the "digital turn" in the new literacy studies. *Review of

Educational Research, 80*, 246-271. doi:10.3102/0034654310364401

National Governors Association Center for Best Practices & Council of Chief State School

Officers. (2010). *Common Core State Standards for English language arts and literacy in

history/social studies, science, and technical subjects*. Washington, DC: Author.

Organisation for Economic Co-operation and Development. (2011). *PISA 2009 results: Students on line. In digital technologies and performance (Vol. 6)*. doi:10.1787/9789264112995-en

Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist, 51*, 59-81. doi:10.1080/00461520.2016.1145550

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Erlbaum.

Rahman, T., & Mislevy, R. J. (2017). *Integrating cognitive views into psychometric models for reading comprehension assessment*. ETS Research Report Series ETS SS-17-35.. Princenton, NJ: Educational Testing Service. doi:10.1002/ets2.12163

Rouet, J. F. (2006). *The skills of document use: From text comprehension to Web-based learning*. Mahwah, NJ: Lawrence Erlbaum.

Rouet, J.-F., Ros, C., Goumi, A., Macedo-Rouet, M., & Dinet, J. (2011). The influence of surface and deep cues on primary and secondary school students' assessment of relevance in Web menus. *Learning and Instruction, 21*, 205-219. doi:10.1016/j.learninstruc.2010.02.007

Sabatini, J. P., O'Reilly, T., Halderman, L. K., & Bruce, K. (2014). Integrating scenario-based and component reading skill measures to understand the reading behavior of struggling readers. *Learning Disabilities Research & Practice, 29*(1), 36-43. doi:10.1111/ldrp.12028

Schacter, J., Chung, G. K., & Dorr, A. (1998). Children's Internet searching on complex problems: Performance and process analyses. *Journal of the Association for Information Science and Technology, 49*, 840-849. doi:10.1002/(SICI)1097-4571(199807)49:9<840::AID-ASI9>3.0.CO;2-D

Smith, B. E., Kiili, C., & Kauppinen, M. (2016). Transmediating argumentation: Students

composing across written essays and digital videos in higher education. *Computers &*

Education, 102, 138-151.

Snow, C. E. (2002). *Reading for understanding: Toward a research and development*

program in reading comprehension. Santa Monica, CA: RAND Corporation.

Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading*

Research Quarterly, 24, 7-26.

Strømsø, H. I., Bråten, I., & Britt, M. A. (2010). Reading multiple texts about climate

change: The relationship between memory for sources and text comprehension. *Learning*

and Instruction, 20, 192-204. doi:10.1016/j.learninstruc.2009.02.001

Tseng, S., & Fogg, B. J. (1999). Credibility and computing technology. *Communications of*

the ACM, 42(5), 39-44.

Wineburg, S., McGrew, S., Breakstone, J., & Ortega, T. (2016). *Evaluating information: The*

cornerstone of civic online reasoning. Retrieved from

https://sheg.stanford.edu/upload/V3LessonPlans/Executive%20Summary%2011.21.16.pdf

Zhang, S., & Duke, N. K. (2008). Strategies for Internet reading with different reading

purposes: A descriptive study of twelve good Internet readers. *Journal of Literacy*
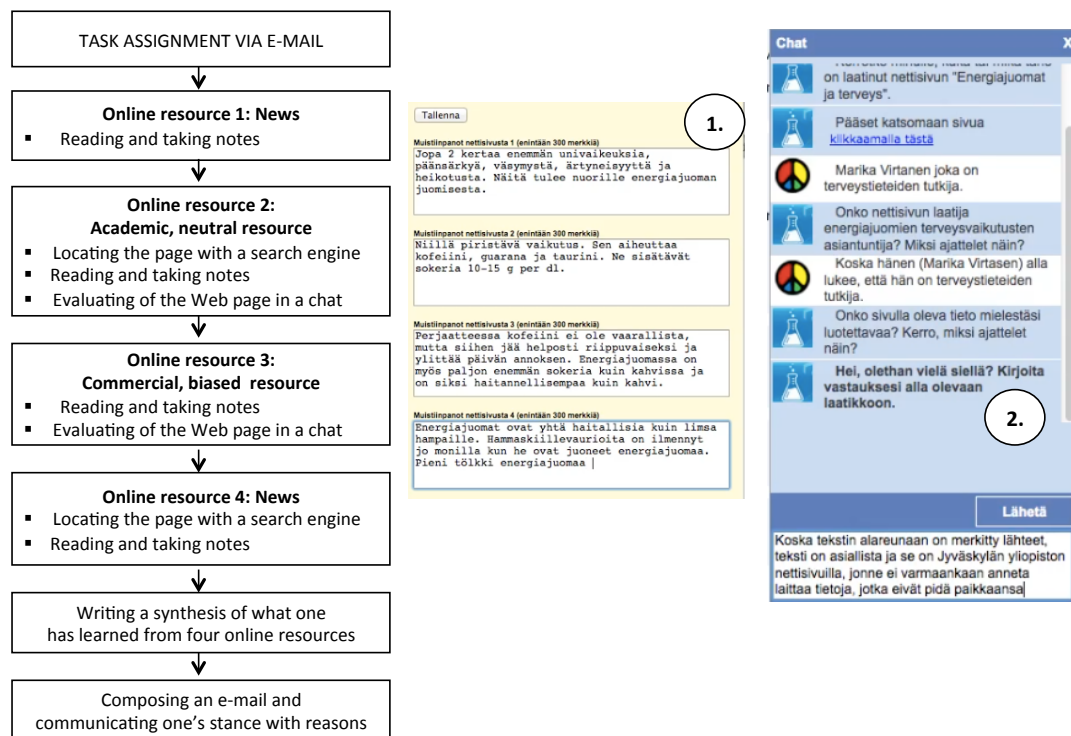
Figure 1. Flow of the online research and comprehension task, and screen shots of (1) the note-taking tool, and (2) the evaluation chat environment
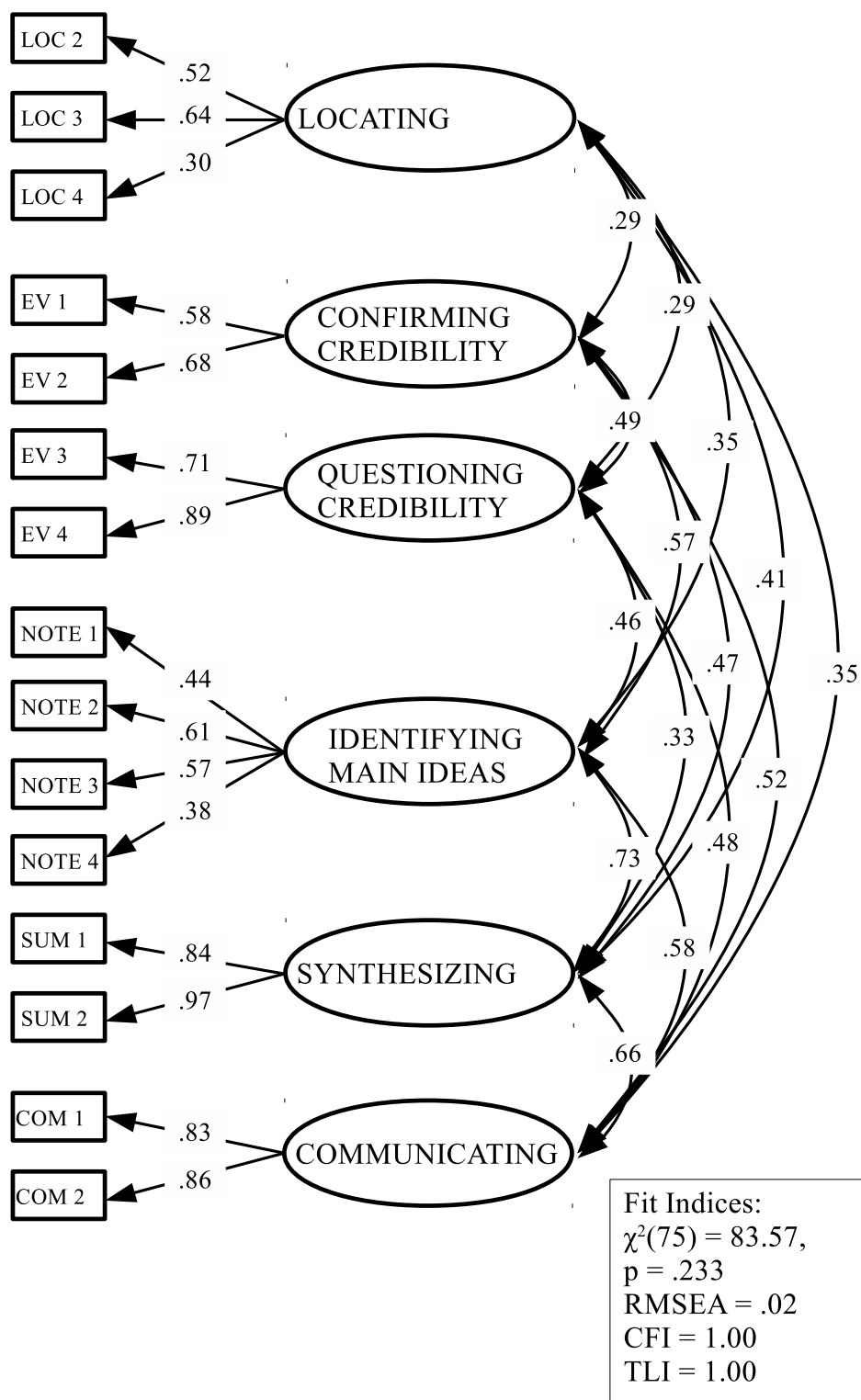
Figure 2. Six-factor measurement model of online research and comprehension including standardized parameter estimates and fit indices.

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis index.

Figure 3. Four-factor measurement model of online research and comprehension, including

standardized parameter estimates and fit indices.

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis index.

Figure 4. Five-factor measurement model of online research and comprehension, including

standardized parameter estimates and fit indices.

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis index.
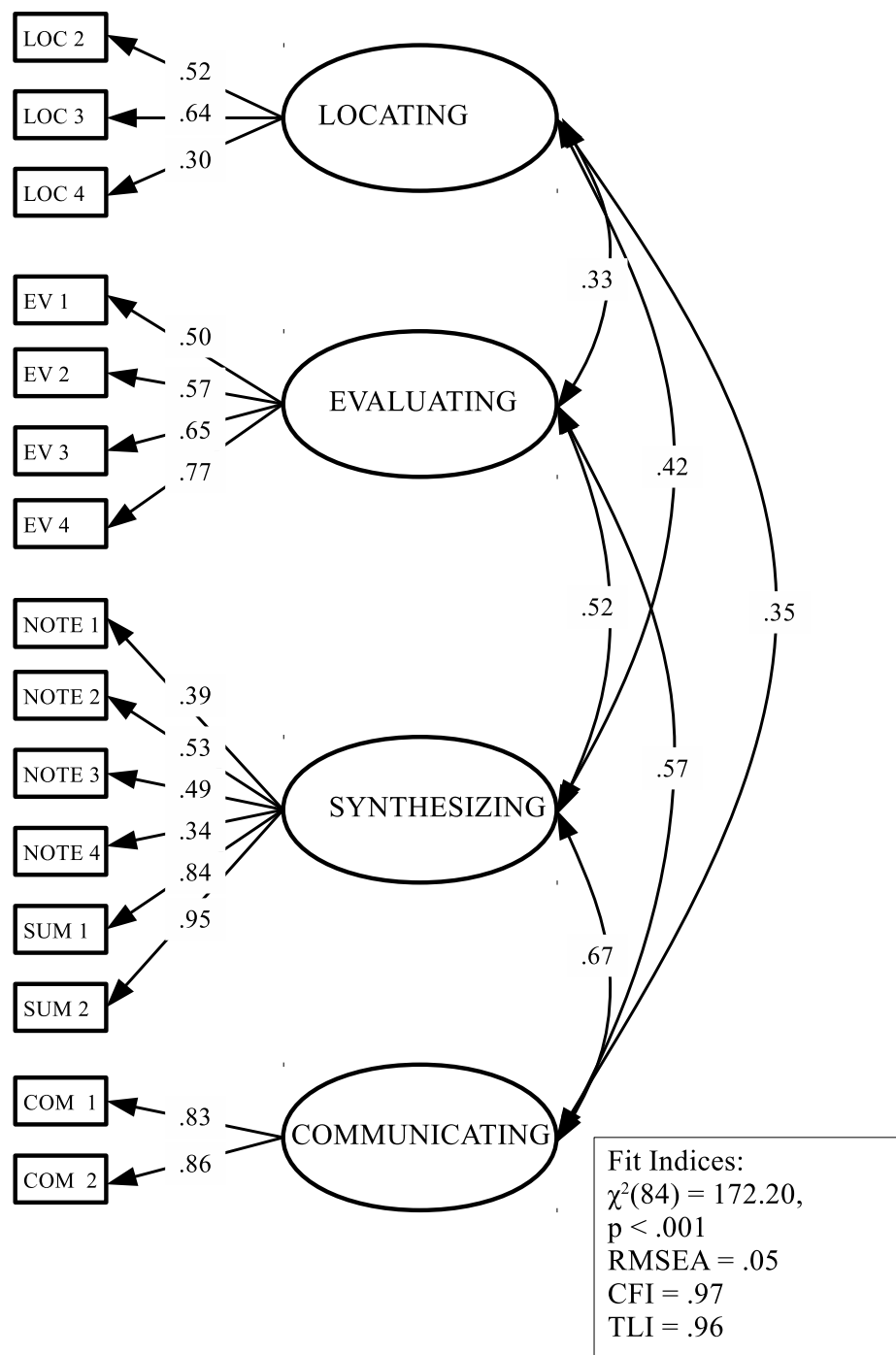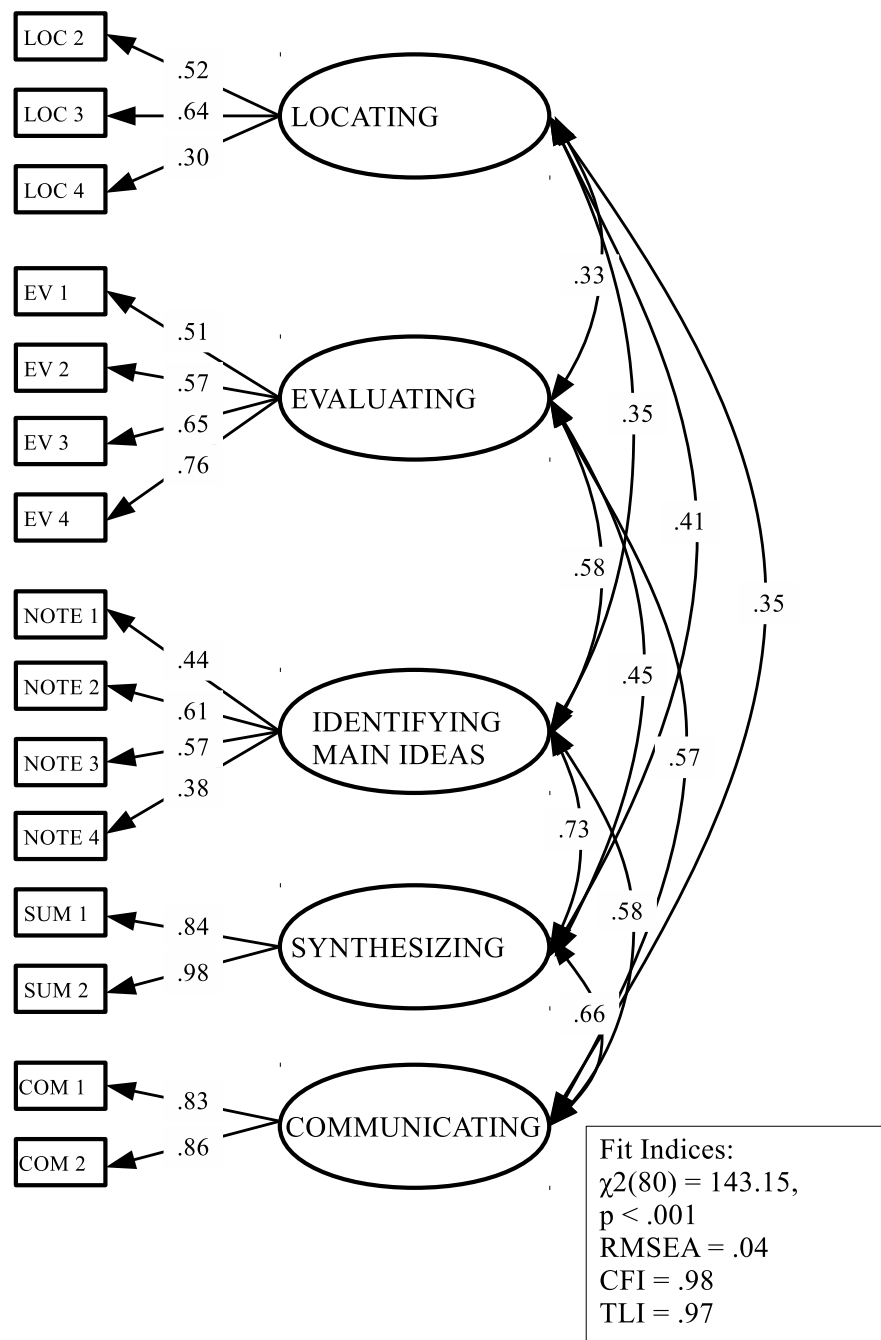
Table 1

The Characteristics of the Online Resources Students Read in the Research Task and Processes

That Were Related to Them (Given/Located; Evaluated)

Title	Genre	Description	Given/ Located	Evaluated
1. Energy drinks are associated with adolescents' sleeping disorders	News	The news article reports the results of a large national study concerning the effects of energy drinks on adolescents.	Given	
2. Energy drinks and health	Informational	A researcher from a health department answers parents' common questions concerning energy drinks. The source provides facts on health issues related to the use of energy drinks without taking a position either for or against their use. Rather, it informs readers about the circumstances when energy drinks are considered safe to use and when they are not. At the end of the page, the references used are listed.	Located	Yes
3. New energy drink to the stores	Press release	This Web page of an imaginary manufacturer has the primary purpose of promoting sales of a new energy drink product. Argumentation included in the source is one-sided as the head of the marketing department takes a positive stance toward the health effects of energy drinks. Only at the last line is slight criticism related to energy drinks expressed: "However, too much is too much."	Given	Yes
4. There are 14 lumps of sugar in an energy drink can	News	In the news article, a researcher from the National Institute for Health and Welfare indicates her concern about the increasing consumption of energy drinks by youth. The researcher tells about the consequences of energy drinks on health, including the health of the teeth.	Located	No

Table 2

Scoring Criteria for Online Research and Comprehension Components

Component	Measured Aspect	Description	Number of Items x Scores	Inter-Rater Reliability
Locating	Formulation of a search query	Ability to use relevant search terms in the query	2 x 0–2 p.	Query 1: 1.000 Query 2: 1.000
	Search efficiency: Time spent on locating the correct online resource	Ability to efficiently select the relevant link from the irrelevant ones offered by the search engine	2 x 0–4 p.	Not applicable
Evaluating	Evaluation of authors' expertise	Ability to evaluate authors' expertise on health issues when reading either an academic, neutral resource (AR) or commercial, biased resource (CR) online	2 x 0–3 p.	AR: 0.968 CR: 0.950
	Evaluation of credibility of information	Ability to evaluate credibility of information when reading either an academic, neutral or commercial, biased resource	2 x 0–3 p.	AR: 0.947 CR: 0.983
Synthesizing	Identifying main ideas from a single online resource	Ability to gather main ideas indicating either the relative nature (online resources 1–3) or causality (online resource 4) of health effects of energy drinks	4 x 0–2 p.	Note 1: 1.000 Note 2: 0.916 Note 3: 0.858 Note 4: 0.870
	Utilizing multiple resources in the summary	Ability to utilize multiple online resources in the summary	1 x 0–3 p.	0.81
	Integration of ideas in the summary	Ability to compose a versatile and coherently structured summary	1 x 0–3p.	0.784
Communicating	Quality of argumentation in the e-mail	Ability to take a stance that the online resources support and to present multiple supporting reasons representing different perspectives	1 x 0–5 p.	0.939

| Communicative practices in the e-mail | Ability to compose an e-mail message that communicates multiple ideas clearly and politely, showing awareness of the audience | 1 x 0–5 p. | 0.722 |

Table 3

Scoring Criteria for Locating, Evaluating, Synthesizing, and Communicating Items

---

Locating

Formulation of search query

2 points	Query contains the topic (energy drinks) + focus (University of Jyväskylä in the first search task and teeth in the second search task)
1 point	Query contains only topic OR topic + focus + at the most one additional word
0 points	Query does not meet the criteria above (e.g., contains only focus)

Efficiency of locating the correct online resource

4 points	Fast: ≤ 10 s
3 points	Quite fast: 11─30 s
2 points	Quite slow 31─90 s
1 point	Slow 91 ≥
0 point	Student does not find the correct resource.

Evaluating

Evaluation of expertise in an academic, neutral resource

3 points	Response explicitly or implicitly states that the author is an expert and justifies his or her statement by referring to credibility labels (e.g., the author's research background, organization, or expertise on the topic) that are presented on the online resource.
2 points	Response explicitly or implicitly states that the author is an expert and student also justifies his or her statement with relevant reason(s) other than the credibility labels presented on the online resource (e.g., knows a lot, talks like an expert).
1 point	Response states that the author is an expert but does not justify his or her statement in a relevant way.
0 point	Response does not explicitly or implicitly state that the author is an expert.

Evaluation of overall credibility of an academic, neutral resource

3 points	Response explicitly or implicitly states that the online resource is credible and justifies his or her statement with at least two relevant reasons.
2 points	Response explicitly or implicitly states that the online resource is credible and justifies his or her statement with one relevant reason.
1 point	Response states that the online resource is credible but does not justify his or her statement.
0 point	Response does not evaluate the online resource or evaluates it as not credible.

Evaluation of expertise in a commercial, biased resource

3 points	Response explicitly or implicitly questions the expertise of the author by noting that the author's expertise is related to some area other than health issues (e.g., sales promotion or as a representative of the company).

2 points	Response explicitly or implicitly questions the expertise of the author by referring to aspects of the information that the author provides or does not provide (e.g., the author does not tell about the negative effects of energy drinks, or she provides information that is inconsistent with that in other sources).
1 point	Response questions the expertise of the author but does not justify his or her judgment in a relevant way.
0 point	Response does not include any doubt about the expertise of the author.

Evaluation of overall credibility of commercial, biased resource

3 points	Response questions the credibility of the online resource and justifies his or her judgment on the basis of commercial bias.
2 points	Response questions the credibility of the online resource. The student does not recognize the commercial bias but justifies his or her critical judgment with some other relevant reason(s) (e.g., one-sided information, information that is inconsistent with one's previous knowledge).
1 point	Response questions the credibility of the online resource. The student does not justify his or her judgment in a relevant way.
0 point	Response does not include any doubt about the credibility of the online resource.

Synthesizing

Note 1

2 points	Note includes the idea that energy drinks may cause multiple symptoms AND that the effects may differ between girls and boys.
1 points	Note includes the idea that energy drinks may cause multiple symptoms OR may differentially affect girls and boys.
0 points	Note does not fulfill criteria required to earn 1 or 2 points.

Note 2

2 points	Note includes the idea that effects of energy drinks are dependent upon age, size of a person, or amount of use (at least one of these aspects required).
1 points	Note includes at least one of the relevant facts related to the health effects of energy drinks.
0 points	Note does not fulfill criteria required to earn 1 or 2 points.

Note 3

2 points	Note includes both sides of the following aspects: moderate consumption of energy drinks is not harmful (or can even have some positive effects) but heavy consumption is harmful.
1 points	Note includes only one side of the following aspects: moderate consumption of energy drinks is not harmful (or can even have some positive effects) or heavy consumption is harmful.
0 points	Note does not fulfill criteria required to earn 1 or 2 points.

Note 4

2 points   Note explicitly includes the causal connection between energy drinks (or sugar included in energy drinks) and damage in tooth enamel.

1 points   Note includes relevant issues concerning energy drinks and health of the teeth, but it does not explicitly state the causal connection between sugary energy drinks and tooth damage.

0 points   Note does not fulfill criteria required to earn 1 or 2 points.

Utilizing multiples sources in the summary

3 points   Four Web pages are used in response.

2 points   Three Web pages are used in response.

1 point   Two Web pages are used in response.

0 points   Only one Web page or none of the Web pages are used in response.

Quality of integration of ideas in the summary

3 points   Summary is versatile and fully coherent.

2 points   Summary is versatile, but not fully coherent (e.g., contains some elements of coherence such as connective ties) OR summary is coherent, but only somewhat versatile.

1 point   Summary lists several health issues about energy drinks but ideas are disconnected in their presentation OR summary contains only a few health issues of energy drinks but contains some elements of coherence.

0 points   Summary is very limited.

Communicating

Quality of argumentation

5 points   Student stands against purchasing energy drink vending machine and gives at least 10 supporting reasons that represent more than five different perspectives (e.g., physical health, mental health, dental health, sleep, schoolwork).

4 points   Student stands against purchasing energy drink machine and gives at least seven supporting arguments that represent more than three different perspectives.

3 points   Student stands against purchasing energy drink machine and gives at least four supporting arguments that represent more than two different perspectives.

2 points   Student stands in support of purchasing energy drink machine in certain conditions and gives 2 or 3 supporting arguments for this stance OR student has not expressed a stance but gives at least five arguments that implicitly support a certain stance.

1 point   Student stands in support of purchasing energy drink machine in certain conditions and gives one supporting argument for his or her stance.

0 points   Students' e-mail does not fulfill criteria required to earn 1 to 5 points.

Communication practices in the e-mail

5 points   E-mail includes at least a somewhat informative title and message content is informative (at least five relevant, distinctive ideas). The message is polite, contains no errors in capitalization and punctuation, and shows awareness of an audience.

4 points	E-mail includes at least a somewhat informative title and message content is somewhat informative (at least three distinctive ideas). The message contains only a few errors in capitalization or punctuation, and it includes a greeting in the beginning of the message or shows some other type of awareness of an audience.
3 points	E-mail includes at least a somewhat informative title and message content is somewhat informative (at least three distinctive ideas). The message contains only a few errors in capitalization or punctuation.
2 points	E-mail fulfills two of the following criteria: (1) includes a title that is at least somewhat informative, (2) the content of the message is quite informative, and (3) the message includes only a few errors in capitalization or punctuation.
1 point	E-mail is not rude, but it is very short or its informativeness is weak, or the content is unclear or poorly written.
0 points	E-mail or its title is rude.