



# Assessing University Students' Critical Online Reasoning Ability: A Conceptual and Assessment Framework With Preliminary Evidence

Dimitri Molerov<sup>1\*</sup>, Olga Zlatkin-Troitschanskaia<sup>2</sup>, Marie-Theres Nagel<sup>2</sup>, Sebastian Brückner<sup>2</sup>, Susanne Schmidt<sup>2</sup> and Richard J. Shavelson<sup>3</sup>

<sup>1</sup> Department of Research Methods in Education, Humboldt University of Berlin, Berlin, Germany, <sup>2</sup> Department of Business and Economics Education, Johannes Gutenberg University, Mainz, Germany, <sup>3</sup> Stanford Graduate School of Education, Stanford University, Palo Alto, CA, United States

## OPEN ACCESS

### Edited by:

Douglas F. Kauffman,  
Medical University of the  
Americas – Nevis, United States

### Reviewed by:

Henk Huijser,  
Queensland University of  
Technology, Australia  
Ronny Scherer,  
University of Oslo, Norway

### \*Correspondence:

Dimitri Molerov  
molerov@hu-berlin.de

### Specialty section:

This article was submitted to  
Educational Psychology,  
a section of the journal  
Frontiers in Education

**Received:** 30 June 2020

**Accepted:** 13 November 2020

**Published:** 15 December 2020

### Citation:

Molerov D, Zlatkin-Troitschanskaia O,  
Nagel M-T, Brückner S, Schmidt S  
and Shavelson RJ (2020) Assessing  
University Students' Critical Online  
Reasoning Ability: A Conceptual and  
Assessment Framework With  
Preliminary Evidence.  
Front. Educ. 5:577843.  
doi: 10.3389/feduc.2020.577843

Critical evaluation skills when using online information are considered important in many research and education frameworks; critical thinking and information literacy are cited as key twenty-first century skills for students. Higher education may play a special role in promoting students' skills in critically evaluating (online) sources. Today, higher education students are more likely to use the Internet instead of offline sources such as textbooks when studying for exams. However, far from being a value-neutral, curated learning environment, the Internet poses various challenges, including a large amount of incomplete, contradictory, erroneous, and biased information. With low barriers to online publication, the responsibility to access, select, process, and use suitable relevant and trustworthy information rests with the (self-directed) learner. Despite the central importance of critically evaluating online information, its assessment in higher education is still an emerging field. In this paper, we present a newly developed theoretical-conceptual framework for Critical Online Reasoning (COR), situated in relation to prior approaches ("information problem-solving," "multiple-source comprehension," "web credibility," "informal argumentation," "critical thinking"), along with an evidence-centered assessment framework and its preliminary validation. In 2016, the Stanford History Education Group developed and validated the assessment of *Civic Online Reasoning* for the United States. At the college level, this assessment holistically measures students' web searches and evaluation of online information using open Internet searches and real websites. Our initial adaptation and validation indicated a need to further develop the construct and assessment framework for evaluating higher education students in Germany across disciplines over their course of studies. Based on our literature review and prior analyses, we classified COR abilities into three uniquely combined facets: (i) online information acquisition, (ii) critical information evaluation, and (iii) reasoning based on evidence, argumentation, and synthesis. We modeled COR ability from a behavior, content, process, and development perspective, specifying scoring rubrics in an evidence-centered design. Preliminary validation results from expert

interviews and content analysis indicated that the assessment covers typical online media and challenges for higher education students in Germany and contains cues to tap modeled COR abilities. We close with a discussion of ongoing research and potentials for future development.

**Keywords:** critical online reasoning assessment, critical thinking, web credibility, higher education, information problem solving using the Internet, multiple-source use, test validation, performance assessment

## INTRODUCTION

### Relevance and Research Background

Today, higher education students use the Internet to access information and sources for learning much more frequently than offline sources such as textbooks (Gasser et al., 2012; Maurer et al., 2020). However, there have been warnings about the harmful effects of online media use on students' learning (Maurer et al., 2018), with misinformation and the acquisition of (domain-specific) misconceptions and erroneous knowledge being prominent examples (Bayer et al., 2019; Center for Humane Technology, 2019). While Internet users are generally concerned about their ability to distinguish warranted, fact-based knowledge from misinformation<sup>1</sup> (Newman et al., 2019), research on web credibility suggests that Internet users pay little attention to cues indicating erroneous information and a lack of trustworthiness; similar findings were determined across a variety of online information environments and learner groups (Fogg et al., 2003; Metzger and Flanagin, 2013, 2015).

For learning in higher education, the Internet may have both a positive and a negative impact (Maurer et al., 2018, 2020). Positive affordances for collaboration, organization, aggregation, presentation, and the ubiquitous accessibility of information have been discussed in research on online and multimedia learning (Mayer, 2009). However, problems such as addictive gratification mechanisms, filter bubbles and algorithm-amplified polarization, political and commercial targeting based on online behavior profiles, censorship, and misinformation (Bayer et al., 2019; Center for Humane Technology, 2019) have recently been critically discussed as well. The potential of online applications and social media for purposes of persuasion has been known for some time (Fogg, 2003), though the impact of online information on knowledge acquisition is still under-researched to date.

As recent research indicates, the multitude of information and sources available online may lead to information overload (Batista and Marques, 2017; Hahnel et al., 2019). Lower barriers to publication and the lack of requirements for quality assurance, fewer gatekeepers, and faster distribution result in a highly diverse online media landscape and varying information quality (Shao et al., 2017). Students are confronted with quality shortcomings such as incomplete, contradictory, or erroneous information when obtaining and integrating new information

from multiple online sources (List and Alexander, 2017; Braasch et al., 2018). Hence, whenever Internet users are acquiring knowledge based on online information or performing online search queries in a way that can be framed as solving an information problem (Brand-Gruwel et al., 2005), they are faced with the challenge of finding, selecting, accessing, and using suitable information. In addition, online learners need to avoid distractions (e.g., advertisements, clickbait) and misinformation as well as evaluate the information they choose with regard to possible biases and specific narrative framing of information (Walton, 2017; Banerjee et al., 2020). To successfully distinguish between trustworthy and untrustworthy online information, students need to judge its relevance to their inquiry and, in particular, evaluate its credibility (Flanagin et al., 2010; Goldman and Brand-Gruwel, 2018). The ability to find suitable information online, distinguish trustworthy from untrustworthy information, and reason based on this information is examined under the term of "critical online reasoning." These abilities are crucial for (self-)regulated (unsupervised) acquisition of warranted (domain-specific) knowledge based on online information.<sup>2</sup> In this context, current studies are focusing on the development of (domain-specific) misconceptions and the acquisition of erroneous knowledge over the course of higher education studies, specifically among students who report that they predominantly use Internet sources when studying (Maurer et al., 2018, 2020).

### University Students' Critical Online Reasoning Assessment (CORA): Study Context

To acquire reliable and warranted (domain-specific) knowledge, students need to access, evaluate, select, and ultimately reason based on relevant and trustworthy information from online sources. At the same time, they need to recognize erroneous or (intentionally) misleading information and possible corresponding bias, for instance, due to underlying framing or unwarranted perspectives, to avoid being misled and acquiring erroneous knowledge. To properly handle online sources featuring incorrect, incomplete, and contradictory information, students need to recognize patterns in the information indicating its trustworthiness or lack thereof (cues for credibility or misinformation) based on self-selected criteria such as perceived

<sup>1</sup>In this study, we focused on misinformation. Misinformation may result from (often unintentional) error, lacking quality assurance, and lacking truth commitment, while disinformation may be spread purposefully due to vested (e.g., business-related, political, ideological, and potentially hidden) interests of stakeholders (Metzger, 2007; Karlova and Fisher, 2013).

<sup>2</sup>We focused on inquiry-based learning using the Internet, information problem solving, and integration of information from multiple sources (Zhang and Duke, 2008; List and Alexander, 2017) in the context of university studies, although the critical evaluation of information when acquiring knowledge while using the Internet for other purposes, such as for entertainment, is important as well.

expertise or communicative intentions to acquire reliable, warranted (domain-specific) knowledge using the Internet.

Students' critical evaluation skills when dealing with online information are considered important in many research frameworks in a multitude of disciplines that address the online learning-and-teaching environment (Section Theoretical and Conceptual Framework; **Table 1**). Like critical thinking and information literacy, they are considered to be among the key twenty-first century skills, and are considered key skills for "Education in the Digital World" (National Research Council, 2012; KMK, 2016). Skills related to the critical-reflective use of online information are more important than ever, which becomes evident especially with regard to the internet-savvy younger generations (Wineburg et al., 2018). Higher education can play a special role in promoting students' critical thinking skills and their skills in evaluating (online) sources (Moore, 2013) due to the evidence-based, research-focused orientation of most academic disciplines (Pellegrino, 2017). For instance, graduate students were found to have advanced critical thinking skills, which has been attributed to the fact that they wrote a bachelor thesis as part of their undergraduate studies (Shavelson et al., 2019; Zlatkin-Troitschanskaia et al., 2019).

Despite being of central importance for studying using the Internet, the assessment of students' skills related to critical online reasoning (COR) is an emerging field with conceptual and theoretical frameworks building on a large number of prior research strands (Section Theoretical and Conceptual Framework; **Table 1**). For instance, computer skills, digital and information literacy, and critical thinking approaches have described and examined (bundles of) related facets. To our knowledge, there is no conceptual and assessment framework to date that describes and operationalizes COR as an interrelated triad of its key facets (i) information acquisition in the online environment, (ii) critical information evaluation, and (iii) reasoning using evidence, argumentation, and synthesis.

In this context, pioneering work has been done by Wineburg et al. (2018) from the Stanford History Education Group (SHEG), who developed an assessment for measuring *Civic Online Reasoning* at the middle school, high school, and college level. At the college level, this holistic assessment of how students evaluate online information and sources comprises short evaluation prompts, real websites, and an open Internet search (Wineburg and McGrew, 2016; Wineburg et al., 2016a,b). The assessment was validated in a nationwide study in the U.S. (Wineburg et al., 2018), which indicated substantial deficits in these skills among higher education students.

Based on this U.S. research, we adapted the assessment framework for higher education in Germany. The preliminary validation of the U.S. assessment for Germany indicated that an adaptation and validation in terms of the recommendations by the international Test Adaptation Guidelines [TAGs, International Test Commission (ITC), 2017] was not possible. It became evident that, in addition to the practical difficulties of adapting the U.S. assessment web stimuli for assessing the critical evaluation of online information for learning in the German higher education context, expert interviews (Section Content Analysis: CORA Task Components as Coverage of the Construct)

indicated that due to the differences in terms of historical and socio-cultural traditions between the two countries, in German higher education, the concept of "civic education" is less prominent than "academic education" (for a comparison of the concept of education/ "Bildung" in Germany and in the U.S., see (Beck, 2020); for a model of critical thinking, see Oser and Biedermann, 2020). Moreover, experts noted that students learn from information from a variety of sources not necessarily related to civic issues (e.g., commercial websites), in addition to scientific publications and textbooks, and it remains unclear how new knowledge based on these multiple sources is integrated, which requires further differentiation and specification.

Based on the results of this preliminary validation, we modified the theoretical framework by expanding our focus beyond civic reasoning to include further purposes of online information acquisition, and situated the construct in relation to a number of theories, models, and adjacent fields, focusing on the research traditions of *critical thinking* (Facione, 1990), which are more applicable to Germany (than civic reasoning), as well as in relation to additional relevant constructs such as "web credibility," "multiple source comprehension," "multiple-source use," and "information problem-solving" using the Internet (Metzger, 2007; Braasch et al., 2018; Goldman and Brand-Gruwel, 2018). Based on a combination of converging aspects from these research strands, we developed a new conceptual framework to describe and operationalize the abovementioned triad of key facets underlying the resulting skill of *Critical Online Reasoning (COR)*: (i) *online information acquisition*, (ii) *critical information evaluation*, and (iii) *reasoning using evidence, argumentation, and synthesis*.

## Research Objectives and Questions

The first objective of this paper is to present this newly developed conceptual and assessment framework, and to locate this conceptualization and operationalization approach in the context of prior and current research while critically reflecting on its scope and limitations. The methodological framework is based on an evidence-centered assessment design (ECD) (Mislevy, 2017). According to ECD, alignment of a *student model*, a *task model*, and an *interpretive model* are needed to design assessments with validity in mind. The *student model* covers the abilities that students are to develop and exhibit (RQ1); the *task model* details how abilities are tapped by assessment tasks (RQ2); and the *interpretive model* describes the way in which scores are considered to relate to student abilities (RQ3). The following research questions (RQ) are examined in this context.

*RQ1: What student abilities and mental processes does the CORA cover? How can the COR ability be described and operationalized in terms of its construct definition?*

*RQ2: What kinds of situations (task prompts), with which psychological stimuli (i.e., test definition), are required to validly measure students' abilities and mental processes in accordance with the construct definition?*

As a second objective of this paper, we focus on the preliminary validation of the COR assessment (hereinafter referred to as CORA). The validation framework for

**TABLE 1** | Theoretical and conceptual background of COR.

<b>Critical online reasoning (COR)</b>		
<b>Main assessment frameworks</b>		<b>Studies (selection)</b>
Civic online reasoning assessment (CORA) (assessment in U.S., employing real websites and live open web search as stimuli)		Wineburg et al., 2016a; Wineburg and McGrew, 2017; Wineburg et al., 2018; Breakstone et al., 2019; McGrew et al., 2019
Performance Assessment of Learning in higher education (PAL) (criterion-sampled performance tasks)		Shavelson et al., 2018, 2019; Zlatkin-Troitschanskaia et al., 2019
Positive Learning in the Age of Information (PLATO) framework		Zlatkin-Troitschanskaia et al., 2018, 2020b
<b>Related research strands for main COR facets</b>		<b>Studies (selection)</b>
Overarching	Multiple-source comprehension and use (MSC)	for an overview, see Rouet, 2006; Lawless et al., 2012; List and Alexander, 2017; Braasch et al., 2018
	Information-problem solving using the Internet (IPS-I)	Brand-Gruwel et al., 2009; Walraven et al., 2009; Goldman and Brand-Gruwel, 2018
	Internet reading strategies/online search behavior and self-reported search strategies	Salmerón et al., 2005; Zhang and Duke, 2008; Zhang et al., 2011; Pernice, 2017
	"Sourcing" in MSC/use of source cues about author and meta-data, for credibility evaluation; proactive, repeated, and task-related sourcing	Braten et al., 2018; Hahnel et al., 2019
(I) Online Information Acquisition (OIA)	Higher education students' use of online information	Head and Eisenberg, 2009; Samson, 2010; Maurer et al., 2018
	Students' information needs and problems using online information and databases	Walraven et al., 2008; Catalano, 2013; Sanders et al., 2015; Kohnen and Mertens, 2019
	Heuristic and systematic information seeking strategies	Chen and Chaiken, 1999; De Neys, 2006; Toplak et al., 2007; Evans and Stanovich, 2013; Gronchi and Giovannelli, 2018
	Information foraging theory	Pirolli and Card, 1999; Juvina and van Oostendorp, 2008
	Interactive Information Retrieval/programming and designing search engines, static and dynamic websites, databases, etc. with a focus on user's interaction with them	for an overview, see (Xie, 2008)
(II) Critical Information Evaluation (CIE)	Disinformation and misinformation classification and current examples (EU)/recent misinformation and disinformation phenomena	(Karlova and Fisher, 2013; Ciampaglia, 2018); for a review of disinformation as a threat to democracy, see also (Bayer et al., 2019; Flore et al., 2019)
	Media bias and propaganda/Studies on strategy and system-level media effects promoting spread of misinformation and disinformation, including the documentation of deconstructions of media, framing, interaction measures, message components, power exertion, and manipulation techniques in advertisement, propaganda, and journalism, as well as implications for (civic) education	Herman and Chomsky, 2002; Paul and Elder, 2008; Daniels, 2009; Walton, 2017
	Hierarchy of Influences on a media message/media-sociological approach highlighting agents and practices influencing media messages at different levels of power, correspondingly, different scope and spread of misinformation	Reese and Shoemaker, 2016
	General Web credibility models	Prominence Interpretation Theory/cue identification and interpretation for operator, design, and content; gullibility and incredulity error; deceiver credibility Two-step judgment/immediate surface judgment and subsequent message judgment

(Continued)

**TABLE 1** | Continued

<b>Critical online reasoning (COR)</b>		<b>Studies (selection)</b>
	Dual processing model/3 phases: ability and motivation at exposure influence propensity and depth of evaluation; follow-up studies on relations and features	Metzger, 2007; Winter et al., 2016; Flanagin et al., 2018; Krämer et al., 2018
	Unifying model/ "construct" phase—subjective and context-dependent criterialization of credibility; heuristics and interaction	Hilligoss and Rieh, 2008
	MAIN model—modality, agency, interactivity, navigability/affordances of technology itself as cues for credibility	Sundar, 2008
	Information trust/3 "s" model (surface, source, semantics), influence by users' domain knowledge, topic knowledge, and information skills	Lucassen and Schraagen, 2011, 2013
	New Web Credibility model/juxtaposes P-I website dimensions (operator, content, design) with credibility attributions (expertise and trustworthiness); overview	Choi, 2015
	Content Credibility Corpus/corpus collection of websites and topics for credibility evaluation	Kakol et al., 2017; Wierzbicki, 2018
Web credibility aspects (selection)		appearance (Akamine et al., 2008); web 1.0 to 2.0 (Tanaka, 2009; Tanaka et al., 2010); web experience (Jozsa et al., 2012); conflicting topics (Salmerón et al., 2013), fear appeals (Dunbar et al., 2014), relations with trust in press (Go et al., 2016), message sidedness (Flanagin et al., 2018), source credibility in political communication (Flanagin and Metzger, 2014)
Further key evaluation criteria of online information, see also web credibility studies		text relevance (McCrudden et al., 2011); accessibility/comprehensibility (Snow, 2002; Coiro, 2003); usefulness (Goldman et al., 2013)
(III) Reasoning with Evidence, Argumentation and Synthesis (REAS)	Integrated model/review of CT constructs for higher education and the online environment, integrated model, ambiguity experience as activation	Jahn, 2012; Jahn and Kenner, 2018
	Logic-based CT/CT pioneering approach	Ennis, 1985
	Education- and psychology-based CT/Delphi study on CT components	Facione, 1990
	Development of reflective CT/6-stage theory based on systematicity of reflection/meta-cognitive monitoring	Elder and Paul, 2010
	Psychology-based CT	Halpern, 2014
	Scientific reasoning and argumentation/studies focusing on students' reasoning and argumentation patterns based on scientific evidence, models, and principles	Fischer et al., 2014; Fischer, 2018
	Rational thinking/rational thinking construct and operationalization	Stanovich et al., 2016
	Valid (informal) argumentation/fundamental components of argumentation, construction, and analysis of valid argumentation patterns (schemes of (un)warranted reasoning, critical questions for knowledge elicitation)	Walton, 2006; Walton et al., 2008
	Adequate heuristics/fast and frugal heuristics for ecological rationality	Goldstein and Gigerenzer, 2002
	Suboptimal heuristics and biases	Kahneman et al., 1982
	Reasoning fallacies	Van Eemeren, 2013

(Continued)

TABLE 1 | Continued

Critical online reasoning (COR)		Studies (selection)
(IV) Metacognitive Activation (MCA)—Overarching metacognitive and regulative component, affective and attitudinal aspects	Definition of bias	Walton, 2006
	Metacognitive processes and regulation/ review of meta-cognitive processes to support information search	Blummer and Kenton, 2015
	Affective reactions and uncertainty/6-phase Information Search Process with varying certainty and affective response	Kuhlthau, 1993
	Context-based initial activation	Jahn, 2012
	Activation via discrepancy detection/ambiguity experience/subfacet of Discrepancy-Induced Source Comprehension (D-ISC) model for mid-task activation (attitudinal) Critical thinking dispositions	Jahn, 2012; Braasch and Bråten, 2017  openness to experience etc. Facione (1990) intellectual virtues (Paul and Elder, 2005)
<b>Related research strands for broader COR activity</b>		
Literacies	Digital literacy/media literacy, information literacy, and computer literacy	Koltay, 2011; Bulger et al., 2014; Murray and Pérez, 2014; Sparks et al., 2016
	Information literacy	American Library Association, 2000; Kingsley et al., 2011; Taylor and Dalal, 2014; Sanders et al., 2015; Maurer et al., 2017; Podgornik et al., 2017; McMullin, 2018; Walton et al., 2020
	ICT information and communication technology literacy Media literacy	Zylka et al., 2015 Damico and Panos, 2018; Powers, 2019; Threadgill and Price, 2019
Further Relevant Assessment Frameworks (selection)	Multiple-source comprehension assessment	Lawless et al., 2012
	Assessment of argument evaluation in scientific texts Critical Thinking assessment in higher education	Münchow et al., 2019 Liu et al., 2014
Instructional Approaches (selection)	Civic Online Reasoning instructional intervention	McGrew et al., 2019
	Critical source evaluation for improved search	Leeder and Shah, 2016
	Bad News game/multilingual browser game on use of major media disinformation strategies, based on 'inoculation' approach	Roosenbeek and van der Linden, 2019
	Review of CT interventions	Abrami et al., 2008
	Fostering CT using digital media/review of instructional designs in HE	Jahn, 2012; Jahn and Kenner, 2018

CORA is based on approaches by Messick (1989) and Kane (2012). A qualitative evaluation of the CORA yielded preliminary validity evidence based on a content analysis of the CORA tasks, and interviews with experts in media science, linguistics, and test development (Section Content Analysis: CORA Task Components as Coverage of the Construct). Based on the results of content validation studies conducted according to the Standards for Pedagogical and Psychological Testing (AERA et al., 2014; hereinafter referred to as AERA Standards), the following RQ was investigated:

*RQ3: To what extent does the preliminary evidence support the validity claim that CORA measures the participants'*

*personal construct-relevant abilities in the sense of the defined construct definition?*

In Section Theoretical and Conceptual Framework, we first present the theoretical and conceptual COR framework, also in terms of related research approaches. In Section Assessment Framework of Critical Online Reasoning, we describe the U.S. assessment of civic online reasoning and present our work toward adapting and further developing this approach into an expanded assessment framework and scoring scheme for measuring COR in German higher education. In Section Preliminary Validation, we report on initial results from the preliminary validation studies. In Section Research Perspectives, we close with implications for refining CORA tasks and rubrics

and give an outlook on ongoing further validation studies and analyses using CORA in large-scale assessments.

## THEORETICAL AND CONCEPTUAL FRAMEWORK

In this section, we outline the working construct definition for Critical Online Reasoning (COR) as a basis for the CORA framework. We explain the theoretical components and key considerations used to derive this COR construct definition from related prior approaches and frameworks. COR is modeled from a *process, content, domain, and development* perspective. For brevity, we only describe the key facets and central components and list the most relevant references categorized by (sub)facets in **Figure 1**.

### Construct Definition of Critical Online Reasoning

The working construct definition of COR (*RQ1*) describes the personal abilities of searching, selecting, accessing, processing, and using online information to solve a given problem or build knowledge while critically distinguishing trustworthy from untrustworthy information and reasoning argumentatively based on trustworthy and relevant information from the online environment.

This construct definition focuses on a combination of three overlapping facets: (i) *Online Information Acquisition (OIA)* abilities (for inquiry-based learning and information problem-solving), (ii) *Critical Information Evaluation (CIE)* abilities to analyze online information particularly in terms of its credibility and trustworthiness, and (iii) abilities to use the information for *Reasoning based on Evidence, Argumentation, and Synthesis (REAS)*, weighting (contradictory) arguments and (covert) perspectives, while accounting for possible misinformation and biases. In addition, we assume that the activation of these COR facets requires metacognitive skills, described in the *Metacognitive Activation (MCA)* (**Figure 1**).

### Theoretical Components of COR Process Perspective

*Online Information Acquisition (OIA)* focuses on the searching and accessing of online information, for example by using general and specialized search engines and databases, specifying search queries, opening specific websites. Beyond these more technical aspects, COR focuses in particular on searching for specific platform entries and passages and terms on a website in as far as they contribute to an (*efficient*) *accessing of relevant and trustworthy information and avoidance of untrustworthy information* (Braten et al., 2018; the Information Search Process model, Kuhlthau et al., 2008).

*Critical Information Evaluation (CIE)* is crucial for self-directed, cross-sectional learning based on online information. This facet focuses on students' *selection of information sources and evaluation of information and sources* based on website features or specific cues (e.g., text, graphics, audio-visuals). Following comprehension-oriented reception and processing,

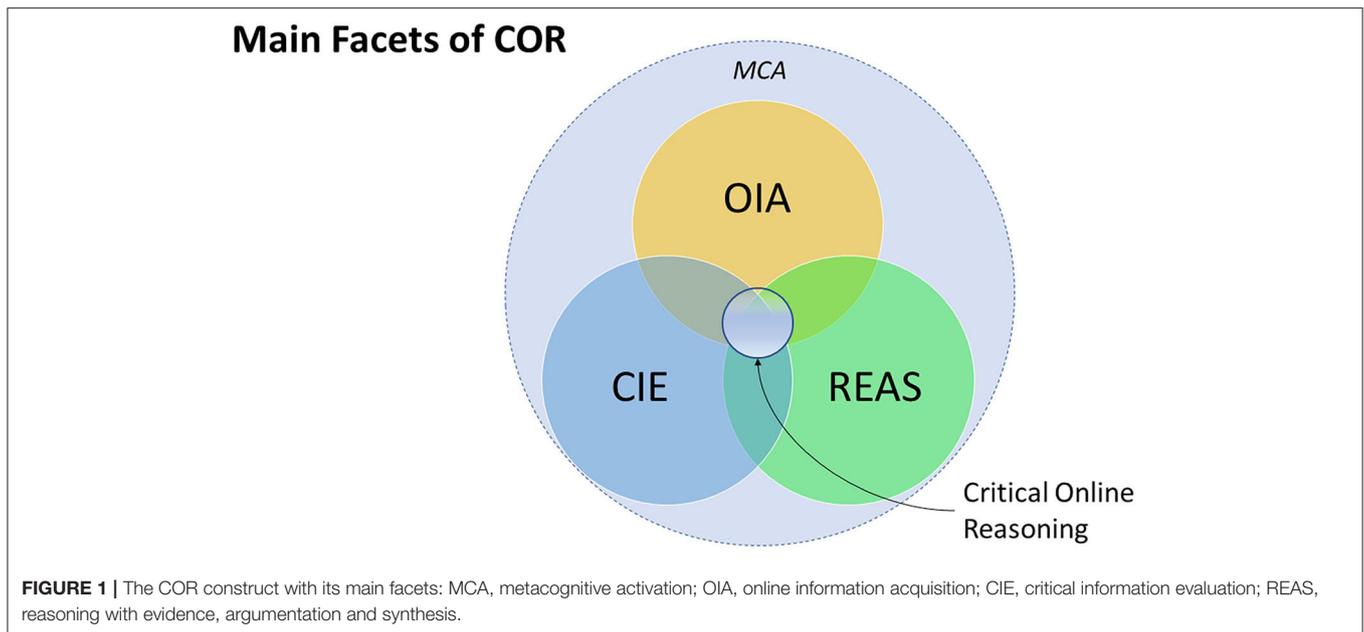
CIE is used to differentiate and select high- instead of low-quality information (relative to one's subjective standards and interpretation of task requirements). A cue can be any meaningful pattern in the online environment interpreted as an indicator of (trustworthy or untrustworthy) online media or communicative means. Examples of cues may be a URL, title or keyword on the search engine results page, a layout or design element, media properties, an article title, information about author, publisher or founder, publication date, certain phrasings, legal or technical information. Trustworthiness "evaluations" typically include targeted verification behavior, which results in a (defeasible) "judgment" about a web medium or piece of information, which may be based on an initial heuristic appraisal without further (re-)evaluation. However, CIE as "evaluation" can require a more systematic analytical, criteria-based judgment process for students, possibly using multiple searches to establish reliable and warranted knowledge (for an overview on related multiple document comprehension frameworks, see Braten et al., 2018; e.g., the Discrepancy-Induced Source Comprehension (DISC) model, Braasch and Bråten, 2017).

*Reasoning with Evidence, Argumentation, and Synthesis (REAS)* is probably the most important facet of COR, which distinguishes this construct from "literacy" constructs (e.g., digital, information or media literacy). This facet focuses on uniting the initially appraised information, weighting it against further indications and perspectives, and using it as evidence to construct a convincing argument that accounts for uncertainty (Walton, 2006). Argumentation is a well-suited discourse format for deliberating whether to accept a proposition (e.g., to trust or distrust). Evidence-based argumentation imposes certain quality standards for a well-founded judgment (e.g., rationality) and requires minimal components of a claim, reasons, evidence (and data) and conventional inferential connections between them (e.g., Argumentation Schemes, Walton, 2006; Walton et al., 2008; Fischer et al., 2014; Fischer, 2018).

These three main facets, OIA, CIE, and REAS, are primarily considered cognitive abilities. Each of them can also take on a metacognitive quality within the COR process, for example as reasoners (internally) comment on their ongoing search, evaluation or argument construction (e.g., "I would not trust this website"), or (self-)reflect on previously acquired knowledge to identify incorrectness or inconsistencies (e.g., "This sentence here contradicts that other source/what I know about the subject"). The latter reflection can become epistemic if it turns to the method of information acquisition and reasoning itself (e.g., "How did I end up believing this scam?").

These main facets are accompanied by an overarching, self-regulative, metacognitive COR component that activates deliberate COR behavior and coordinates (transitions) between the COR facets in the progression of COR activity, particularly for activating a critical evaluation and deciding when to terminate it, in relation to other events (e.g., during a learning experience, social communication)<sup>3</sup>. Self-regulation can be

<sup>3</sup>Regulation of COR may be performed deliberately using meta-cognition, or in response to a cognitive process outcome, habitual behavior, processing of environmental cues, affective or motivational state.



applied to monitor and maintain focus (noticing unfocused processing, returning to task) and handle environmental signals (identifying and minimizing distracting information features) (Blummer and Kenton, 2015). As reasoners may have affective responses (Kuhlthau, 1993) to their task progress and to specific information (particularly on controversial topics), affective self-regulation can not only support them in staying on task and keeping an open mind, but they can use it metacognitively for COR to gain an insight into unconsciously processed information (e.g., identifying and coping with triggered avoidance reactions or anxiety induced by ambiguity or manipulation attempts) and can critically reflect on triggers in the source cues.

Thus, *Metacognitive Activation (MCA)* is assumed to be an ability required to activate COR in relevant contexts. (Epistemic) metacognition can be characterized by gradations of self-awareness regarding information acquisition, evaluation and reasoning processes, which may activate a “vigilance state” in students and lead to certain (subconscious) reactions (and a habitual affective response, e.g., anxiety, excitement), or can also be interpreted as an indicator of a potential problem with processed information (“am I being lied to/at risk after misjudging the information?”) at the metacognitive level (on uncertainty and emotions when searching for information, see Kuhlthau, 1993; on ambiguity experience as the first stage in a general critical reasoning process, see Jahn and Kenner, 2018), which may lead to the activation of an evaluative COR process.

The main facets of COR and the overarching metacognitive self-regulative component are understood to determine *COR performance* (and are the focus of the CORA, Section Test Definition and Operationalization of COR: Design and Characteristics of CORA Tasks). The main COR facets are

assumed to rely on “secondary” sub-facets that provide support in cases where *related specific problems* occur, including self-regulation for minimizing distractions and on-task focus, as well as diverse knowledge sub-facets.

Knowledge sub-facets may include, for OIA, knowledge of resources and techniques for credibility verification; for CIE, knowledge of credibility indicators and potentially misleading contexts and framings, manipulative genres and communication strategies; for REAS, knowledge of reasoning standards as well as fallacies, heuristics, and perceptual, reasoning and memory biases as well as of epistemic limitations for trustworthiness assertions. The list is non-exhaustive, and the knowledge and skills are problem-dependent (e.g., checking for media bias will yield conclusive results only if there is in fact a bias in the stimulus material); they can be expected to impact COR in related cases. Hence, controlling for corresponding stimuli encompassed in the task is recommended.

Attitudinal dispositions for critical reasoning and thinking, such as open-mindedness, fairness, and intellectual autonomy (Facione, 1990; Paul and Elder, 2005) are equally likely candidates for COR influences. These secondary facets are not examined in the current conceptualization.

### Content Perspective

For acquiring information online in a warranted way, students need to successfully identify and use trustworthy sources and information and avoid untrustworthy ones. In contrast, unsuccessful performance is marked by trusting untrustworthy information, a gullibility error, or refusing to accept trustworthy sources, an incredulity error (Tseng and Fogg, 1999). To decide which information to trust and use, students need to judge information in regard to several criteria, including at least the

following: *usefulness*, *accessibility*, *relevance*, and *trustworthiness*. Information may be judged as useful if it advances the inquiry, for instance by supporting the construction of an argument; usefulness may also be understood as a holistic appraisal based on all other criteria. Lack of *accessibility* (or comprehensibility) limits students to the parts of the information landscape that they can confidently access and process (e.g., students may ignore a search result in a foreign language or leave a website with a paywall, but also abandon a text they deem too difficult to locate or understand in the given task time). In an open information environment, successfully judging *relevance* as relatedness or specificity to the topic of inquiry and *trustworthiness* or quality of information enables students to select and spend more time on high-quality sources and avoid untrustworthy sources. Assuming students will attempt to ignore information they judge as untrustworthy, any decision in this regard affects their available information pool for reasoning and learning.

The judgment of trustworthiness as an (inter-)subjective judgment of the objectively verifiable quality of an online media product against an evidential or epistemic standard is central to COR. In more descriptively oriented “web credibility” research, a credibility judgment is understood as a subjective attribution of trust to an online media product; trustworthiness in COR is closely related, but presupposes that the judgment can be based on valid or invalid reasoning (acceptable or unacceptable based on a normative standard) and hence can be evaluated as a skill. Trustworthiness in COR can be considered a warranted credibility judgment. Consequently, COR enables students to distinguish trustworthy from untrustworthy information and, more specifically, various sub-types based on assumed expertise and communicative intent, for example: accidental misinformation due to error, open or hidden bias, deliberate disinformation, and (non-epistemic) “bullshitting.” A more fine-grained judgment is assumed to afford higher certainty, a more precise information selection, and more adequate response to an information problem. To successfully infer the type of information, reasoners may evaluate cues from at least three major strands of evidence about an online medium, including cues on content, logic, and evidence; cues on design, surface structure, and other representational factors; and cues on author, source, funding, and other media production and publication-related factors. Reasoners may evaluate these themselves (using their own judgment), trust the judgment of experts (external judgment), or a combination of the two; when accepting external judgment, instead of the information itself, reasoners need to judge at least their chosen expert’s topic-related expertise and truth-oriented intent.

### Domain-Specificity and Generality

Based on the CORA framework, COR is modeled for generic critical online reasoning (GEN-COR) on tasks and websites that do not require specialized domain knowledge and are suited for young adults after secondary education.

The construct can be specified for study domains (DOM-COR), for instance by defining domain standards of evidence for distinguishing trustworthy from untrustworthy information and typical domain problems regarding the judgment of online information.

### Development Perspective

Different gradations can be derived based on task difficulty, complexity, time, and aspired specificity of reasoning (Sections Test Definition and Operationalization of COR: Design and Characteristics of CORA Tasks and Scoring Rubrics). COR ability levels were distinguished to fit the main construct facets depending on students’ performance in (sub-)tasks tapping OIA, CIE, and REAS (see rubrics in Section Scoring Rubrics; **Table 1**).

## ASSESSMENT FRAMEWORK OF CRITICAL ONLINE REASONING

### Civic Online Reasoning

Wineburg and McGrew (2016) developed an assessment to measure *civic online reasoning*, which they defined as students’ skills in interpreting online news sources and social media posts. The assessment includes real, multimodal websites as information sources (and distractors) as well as open web searches. The construct of civic online reasoning was developed from the construct of news media literacy (Wineburg et al., 2016a). It was conceptualized as a key sub-component of analytic thinking while using online media. The assessment aims to measure whether students are able to competently navigate information online and to distinguish reliable, trustworthy sources and information from biased and manipulative information (Wineburg et al., 2016a).

The students’ skills required to solve the tasks were assessed under realistic conditions for learning using the Internet, i.e., while students performed website evaluations and self-directed open web searches (Wineburg and McGrew, 2017). The computer-based assessment presents students with short tasks containing links to websites with, for instance news articles or social media text and video posts, which students are asked to evaluate. The task prompts require the test-takers to evaluate the credibility of information, and to justify their decision, also citing web sources as evidence. The topics focus on various political and social issues of most US-centric civic interest, typically with conflicting constellations of sources.

Using this assessment, the SHEG surveyed a sample of 7,804 higher education students across the U.S. (Wineburg et al., 2016a), and compared the students’ performance to that of history professors and professional fact checkers. Based on the findings, the search engine results pages designed and implemented an intervention to improve students’ civic online reasoning in higher education (Wineburg and McGrew, 2016; McGrew et al., 2019).

## Critical Online Reasoning Assessment (CORA)

In our project<sup>4</sup>, the initial goal was to adapt this instrument to assess the *civic online reasoning* of students in higher education in Germany and to explore the possibility of using this assessment in cross-national comparisons. The assessment of civic online reasoning features realistic judgment and decision-making scenarios with strong socio-cultural roots, which may engage and tap both the (meta)cognition and the emotional responses of test-takers, as well as their critical evaluation skills. While cultural specificity may present advantages in a within-country assessment, these can become idiosyncratic challenges in cross-national adaptations (e.g., Arffman, 2007; Solano-Flores et al., 2009). Even though we followed the state-of-the-art TAGs by the International Test Commission (ITC) (2017) and the best-practice approach of (Double-)Translation, Reconciliation, Adjudication, Pretesting, and Documentation (TRAPD, Harkness, 2003) in assessment adaptation research (as recommended in the TAGs), after the initial adaptation process (Molerov et al., 2019), both the (construct) definition of *civic online reasoning* and the adapted assessment of civic online reasoning showed limitations when applied to the context of learning based on online information in German academic education. The translation team faced several major practical challenges while adapting the real website stimuli, and the results were less favorably evaluated by adaptation experts. This was a key finding from the adaptation attempts and preliminary construct validation by means of curricular analyses and interviews with experts for German higher education. Both analyses indicated the significant differences in terms of historical and socio-cultural traditions between the higher education systems in the two countries (for details, Zlatkin-Troitschanskaia et al., 2018b). Regarding construct limitations, curricular analysis indicated differences in the relevance of “civic education” within German higher education, highlighting problems for the (longitudinal and cross-disciplinary) assessment of generic abilities in learning based on online information. Expert interviews conducted in the context of adaptation attempts and the preliminary validation of the U.S. conceptual and assessment framework of “civic online reasoning” (for details, Molerov et al., 2019) indicated that the concept of “civic education” is related to a specific research strand of political education and is less important in German higher education than “academic education,” which is more strongly related to research traditions focusing on critical thinking (for a comparison of the concept of education in Germany and in the U.S., see Beck, 2020; Oser and Biedermann, 2020).

Based on this preliminary validation of the U.S. assessment in Germany, we modified the conceptual framework (Section Theoretical Components of COR) to accommodate for the close relationship between COR and generic critical thinking, multiple-source comprehension, scientific reasoning and

informal argumentation approaches (Walton, 2006; Fischer et al., 2014, 2018; Goldman and Brand-Gruwel, 2018; Jahn and Kenner, 2018), and expanded the U.S. assessment framework to cover all online sources that students use for learning. We developed the scoring rubrics accordingly to validly measure the *critical online reasoning* (COR) ability of higher education students of all degree programs in Germany in accordance with our construct definition (Section Construct Definition of Critical Online Reasoning). Thus, new CORA tasks with new scenarios were created to cover the (German) online media landscape used for learning and topics including culturally relevant issues and problems. The assessment framework was expanded to comprise tasks stimulating web searches, the critical evaluation of online information, and students' use of this information in reasoning based on evidence, argumentation and synthesis *to obtain warranted knowledge and solve the given information problems, and to develop coherent and conclusive arguments for their decision* (e.g., draft a short essay or evaluative short report). We also developed and validated the scoring scheme to rate the students' responses to the CORA tasks (Section Scoring Rubrics).

## Test Definition and Operationalization of COR: Design and Characteristics of CORA Tasks

The German CORA project developed a holistic, performance assessment that uses criterion-sampled situations to tap students' real-world decision-making and judgment skills. The tasks/situations merit critical evaluation. Students may encounter such tasks when studying and working in academic and professional domains, as well as in their public and private lives (Davey et al., 2015; Shavelson et al., 2018, 2019). CORA comprises six tasks of 10 min each. CORA is characterized by the use of realistic tasks in a natural online environment (for an example, see **Figure 2**). As tasks are carried out on the Internet, students have an unlimited pool of information from which to search and select suitable sources to verify or refute a claim, while judging and documenting the evidence. Five CORA tasks contain links to websites that may have been published with (covert) commercial or ideological intent, and may, for instance aim to sell products or to convince their audience of a particular point of view by offering low-quality information. The characteristics of the low-quality information offered on websites linked in the CORA tasks included, for instance a selection of information while (intentionally) omitting other perspectives, incorrect or imprecise information, irrelevant and distracting information, and biased framing. The tasks feature snippets of information in online media, such as websites, twitter messages, YouTube videos, put forward by political, financial, religious, media or other groups, some cloaked with covert agendas, others more transparent.

A specific characteristic of the CORA tasks is that only the stimuli and distractors included in the task prompt and the websites linked in the tasks can be manipulated and controlled for by the test developers. Since the task prompt asks the students to evaluate the credibility and trustworthiness of the linked website through a free web search, realistic distractors include,

<sup>4</sup>The German CORA project is part of the cross-university PLATO research program, which examines higher education students' Internet-supported learning for the acquisition of warranted knowledge from various disciplinary perspectives (for an overview, see Zlatkin-Troitschanskaia et al., 2018a; Zlatkin-Troitschanskaia, 2020).

**TASK 1: Vegan protein sources**

Please complete this task within 10 minutes.

Visit the following website:

[URL]

Is this a reliable source of information about vegan protein sources?

---

Note: You can use any information on this website and you can freely search the Internet.

Justify your answer with evidence from the Internet sources used and include the corresponding URLs.

**FIGURE 2** | Example CORA Task prompt (German website).

for instance vividly presented information, a large amount of highly detailed information, (unreferenced) technical, numerical, statistical and graphical data, and alleged (e.g., scientific or political) authority. Depending on the search terms used and the research behavior of the participants, they are confronted with different stimuli and distractors in a free web search, i.e., stimuli and distractors may likely vary significantly from person to person. Thus, while we can control the quality of the websites linked in the CORA tasks, the quality of all other websites that students are confronted with during their Internet research depends solely on their search behavior and can be controlled in the assessment only to a limited extent.

### Stimuli and Distractors of the Linked Websites

Low-quality information on the linked websites can be caused by a lack of expertise of the author(s), belief-related bias, or accidental errors when drawing inferences or citing from other sources. Moreover, the linked sources offer contradictory information or inconsistencies between multiple online texts, which learners need to resolve in the process of acquiring consistent knowledge. In our example (**Figure 2**), the provided link leads to a website that offers information about vegan protein sources. At first glance, the website seems to provide accurate and scientifically sound information about vegan nutrition and protein sources, but upon closer inspection, the information turns out to be biased in favor of vegan protein sources. The article is shaped by a commercial interest, since specific products are advertised. This bias can be noticed by reading the content of the website carefully and critically. The existence of an online shop is another indication of a commercial interest motivating the article. In contrast, the references to scientific studies give a false sense of reliability.

As the construct definition of COR states (Section Construct Definition of Critical Online Reasoning), if students wonder about the trustworthiness of certain online information during the inquiry, this should be a sufficient initial stimulus to activate their COR abilities. Thus, we explicitly include the stimulus at

the beginning of an inquiry task prompt of all CORA tasks. The in-task cues can tap these activation routes even if the students did not respond to the initial prompt at the beginning of the task (**Figure 2**). In the example, the participants are also asked whether the website is reliable to stimulate the COR process and a web search.

Following the ECD (Mislevy, 2017), we describe the *task model* and the *student model* of the CORA in more detail.

### Task Model

Task difficulty in terms of the *cognitive requirements* of the construct dimensions of COR varies through the task properties and the prompt (i.e., difficulty of deciding on a specific solution by considering pros/cons or both). For instance, in the dimension of OIA, task difficulty varies in terms of whether students are required to evaluate a website and related online sources or only a claim and related online sources. The quality of the websites found in the free web searches is likely to significantly vary between test participants, which is not explicitly controlled for in the task and in the scoring of the task performance. This information is only examined in additional process analyses using the recorded log files (Section Analyses of Response Processes and Longitudinal Studies).

In the easy CORA tasks, the web authors were aware that they may be biased and alerted their audience to this fact, for instance by stating their stance directly or by acknowledging their affiliation to a certain position or perspective—the students then had to take these statements into account in their evaluation. In the difficult CORA tasks, the web authors actively tried to conceal the manipulative or biased nature of their published content—and the students had to recognize the techniques these authors employed. In addition, they had to identify the severity of this manipulation and to autonomously decide which information was untrustworthy and should therefore not be taken into consideration. This untrustworthy information can comprise a single word or paragraph, an entire document, all content by a specific author or organization, or even entire platforms (e.g., if

their publication guidelines, practices, and filters allow for low-quality information) or entire geographical areas (e.g., due to biased national discourse).

For each CORA task, we developed a rubric scheme that describes the aforementioned specific features of the websites linked in the task, for instance in terms of credibility and trustworthiness of the information they contained (for details, see Section Scoring Rubrics). To develop the psychological stimuli encountered in CORA tasks (in accordance with the construct definition; Section Construct Definition of Critical Online Reasoning), we based our approach on a specific classification of misinformation by Karlova and Fisher (2013) and on classifications of evaluative criteria of information quality (e.g., topicality, accuracy, trustworthiness, completeness, precision, objectivity) by Arazy and Kopak (2011), Rieh (2010, 2014), and Paul and Elder (2005).

Cues indicating trustworthiness or lack thereof were systematized in evidence strands according to the Information Trust model (Lucassen and Schraagen, 2011, 2013). The model distinguishes evidence on author, content, and presentation, which are aligned with classical routes of persuasion in rhetoric; each requires a different evaluation process. We expand this model by a distinction of personal evaluation vs. trust in a secondary source of information (Table 1).

The *task difficulty level* was gauged in particular by the scope and extent of misinformation based on an adaptation of the Hierarchy of Influences model by Shoemaker and Reese (2014), which assesses agents in the media production process and their relative power to shape the media message—and hence introduce error or manipulation, which need to be judged by students to discern the limits of warranted trust (e.g., at the bottom end are obvious deceptions and errors by the author such as SPAM emails or simple transcription mistakes in a paragraph, while at the top end are high-level secret service operations or a society-wide cultural misconception).

Task difficulty in terms of required argumentative reasoning in CORA was varied in three ways: (1) Scaffolding was added to the task prompts by asking students only for part of the argument (e.g., only pro side, con side, or only specific sub-criteria) to reduce the necessary reasoning steps. (2) The stimuli websites were selected by controlling for (i) scope and (ii) order of bias or misinformation, and for how difficult it is to detect it. *Scope* refers to the comprehensiveness of biases or misinformation based on the adapted Hierarchy of Influences model by Shoemaker and Reese (2014). The *order* is the level of meta-cognition that needs to be assumed in relation to a bias or misinformation. (3) The composition of sources that can be consulted for information (i.e., number of supporting and opposing, or high-quality and low-quality sources) can again be modified only in a closed Internet-like environment (Shavelson et al., 2019; Zlatkin-Troitschanskaia et al., 2019), but it can hardly be controlled for on the open Internet.

The natural online environment used in this assessment constitutes a crucial aspect of the CORA task difficulty (that is also related to the reliable scoring of task performance; Section Scoring Rubrics). In a closed information environment with a finite number of sources, a comprehensive evaluation of all

sources is possible. On the Internet, an indefinitely large number of sources are available. Hence, when solving the CORA tasks, students also need to constantly decide whether to continue examining a selected source to extract more information (and how deeply to process this information, e.g., reading vs. scanning) or whether to attempt to find a more suitable source, and a sample of search hits on a search engine results page, and whether they should use different search terms or even switch to a more specialized search engine or a specific database that might yield more useful information. This aspect is related to the *student model* and the primary aim of students in the context of inquiry-based learning based on online information, which is to gather information to “fill” their knowledge gaps while carrying out a task. Learning in an online environment requires students' initial (and later updated) understanding of the problem in relation to a specific generic or domain-specific task, and recognition of the types of information that are needed to solve a given problem, and then carrying out the steps to locate, access, use, and reason based on information, and finally formulate an evidence-based solution to the problem.

### Student Model

The expected response processes while solving the CORA tasks can be described with a focus on their *basic phases* based on the abovementioned Information Problem-Solving using the Internet (IPS-I) model (Brand-Gruwel et al., 2009): (1) *Defining the information problem*, (2) *Searching information*, (3) *Scanning information*, (4) *Processing Information*, (5) *Organizing and presenting information*. These phases are quite common in many other models and categorizations of information search and also media and digital literacy (e.g., Eisenberg and Berkowitz, 1990; Fisher et al., 2005). For a multi-source information problem, we expect that the processes will be iterated for each new source. An additional meta-cognitive *Regulation* component guides orientation, steering, and evaluation, and can be active throughout interacting at each phase (Brand-Gruwel et al., 2009). Required judgments of information trustworthiness can be situated in the meta-cognitive component of evaluation. Within the evaluative process, trustworthiness judgments might be juxtaposed with judgments of accessibility, relevance and/or usefulness at several points *in addition to the ongoing collection of information for the inquiry*. Based on these categorizations, we developed a fine-grained description of the (sub)processes the students are expected to perform while solving the CORA tasks (Table 2).

In the following, we describe the *student model* with regard to the four main COR facets in more detail.

In CORA, the test takers are required to produce an argumentative conclusive written response based on the consulted and critically evaluated online sources. In line with the older IPS model (Brand-Gruwel et al., 2005), we have added a reflective metacognitive review as an expected process, which may occur at any moment but possibly upon response verification, to highlight that COR may be activated even after an iteration of the IPS-I process or after the whole CORA task has been completed without critical consideration.

**TABLE 2** | Possible and necessary processes contributing to quality criterion judgments, with web content considered, attributed to IPS-I phase (on evaluative review, see CORA-MCA and IPS).

IPS-I phase	Judgments			
	Accessibility	Relevance	Trustworthiness	Usefulness (task coverage)
Define information problem		o projection of possible relevant information		
Search information	x ignoring inaccessible sources (language, no link) (unless useful based on "scent")	o specifying relevant search terms x judging SERP results for task relevance	o considering URL for trustworthiness x specific fact-checking search	
Scan information	o managing accessibility on site or abandoning site (paid service, media activation; perceptual aides)	x overview and attempt to locate relevant information x title and topics match to search purpose o in-site search	o considering credibility cues	x amount of information appraisal
Process information	o difficult language (skim over, research, or abandon source)	o continuous evaluation during reading (on-task monitoring)	o evaluating argument o evaluating identified cues	o post-processing appraisal of coverage (and further steps)
Organize and present information		x selecting information		o on-task monitoring
Evaluative review (anytime)	o reflection on missed sources	o verification of relation to task	o doubt may re-activate COR	o post-task evaluation

x = required/necessarily tapped; o = possible/can manifest.

Judgments of usefulness, accessibility, relevance, and trustworthiness of online information can be attributed to the COR facet CIE that is represented as a (meta-)cognitive evaluating component in the IPS-I model. A judgment may require a more elaborate evaluation based on additional information searches. Hence, we assume that a spontaneous trustworthiness judgment can occur at any stage in the IPS-I model. Additionally, a more deliberate, likely criterion-based, reflective evaluation of information, for instance in terms of its trustworthiness, can be performed as a specific (scheduled) sub-stage—if the student is aware of the need to evaluate the information.

The sections in the IPS-I process for evaluations of trustworthiness and other judgments also indicate that they can be interwoven with comprehension and reasoning activities and with each other (Section Construct Definition of Critical Online Reasoning, **Figure 1**). However, they are also likely to be distributed across several stages and differ in the content of partial evaluations and possible inferences drawn. The more detailed view of judgments and evaluations by search phase indicate that several judgments are likely to occur per phase and judgments of accessibility, relevance, trustworthiness, and usefulness are *differentially important across phases* and touch upon *different sub-questions per phase*. For instance, trustworthiness evaluations can be both fast, if an exclusion criterion is found, or gradual over one or several stages, including the collection of multiple cues. We assume this to be the case for information and sources that students evaluate as part of the CORA task. For other additional sources found during web searches, it is likely the student will evaluate trustworthiness just once and with little effort, i.e., heuristically, if they know

they can go back to searching a more trustworthy source faster. (i.e., it is not the student's intention to find and determine every untrustworthy source on the Internet, but find one that is not untrustworthy and meets their needs). We therefore expect that the CORA tasks tap a judgment of trustworthiness, including either a systematic criterion-based evaluation (to the extent to which the test-taker is aware of criteria for trustworthy or untrustworthy information) and/or a vigilant recognition of the specific information features that may help the participants identify bias and misinformation.

In this context, the Information Search Process model (Kuhlthau, 1993) links behavior, cognition, and affective responses, with cognition being characterized by gradations of *self-awareness* regarding information search process (**Figure 1**). Here, again, we assume multifold interrelations with the metacognitive facet of COR. Therefore, we expect that recognizing a cue in the linked information in the CORA task (i.e., stimuli), indicating possible bias or misinformation, may activate a "vigilance state" in the students and lead to certain (subconscious) reactions (and a habitual affective response, e.g., anxiety, excitement) or can also be interpreted as an indicator of a potential problem ("am I being lied to?/at risk after misjudging the information?") at the metacognitive level, which may lead to the activation of the facet of COR, i.e., (meta)cognition for critical reasoning activation (on the role of uncertainty and emotions when searching for information, see Kuhlthau, 1993). In this regard, we consider the ambiguity experience as the initial stage in a general critical reasoning and evaluation process, i.e., a cognitive appraisal marked by uncertainty about the validity of one's interpretation of the current situation that leads to a need for more clarity (or to avoid the problem-solving situation, e.g.,

in case of a low self-efficacy), which may prompt an expected response behavior during the CORA tasks, i.e., critical reflection and evaluation.

In terms of the *task model*, this ambiguity is tapped by the CORA task description and the prompt, which explicitly asks students to judge the trustworthiness of a given website or claim. Thus, the task prompt is the initial stimulus for students to activate their trustworthiness evaluation since the question of whether or not information is trustworthy is explicitly given by the task prompt; the second are the cues offered by the stimulus materials embedded in the CORA tasks; the third is the reminder in the response field of the CORA task to formulate a short statement to the task questions and to list the consulted online sources (Section Scoring Rubrics). The CORA task prompts explicitly require students to formulate a response, justify it with reasons and arguments, and back these up by citing URLs of sources used to reach their decision. Thus, students' responses comprise the fundamental components of argumentative reasoning (Section Theoretical Components of COR). In CORA, we framed the trustworthiness evaluation through an *argumentative model*, and modeled (possible) stances on a (trustworthiness) issue and their supporting reasons and evidence (cues). Alternatively, students might not reason deeply about it, but apply cognitive heuristics (Kahneman et al., 1982; Metzger and Flanagin, 2013). However, given that it is explicitly prompted in the task, we expect students to apply argumentative reasoning and to be able to identify cognitive heuristics (e.g., authority biases) within their argumentation.

In this context, one aspect is particularly important in terms of the *interpretative model* (Section Scoring Rubrics). Assuming that cognitive biases (e.g., confirmation bias), and motivated reasoning can be tapped by controversial topics as presented in the CORA tasks, an opposing stance toward a given topic (i.e., skeptical) affords more stimuli to be critical and motivates the student to find evidence of misinformation. This is why a balanced selection of various topics was established in CORA. We assume students' initial personal stance on the task and its topic will depend on a number of influences, controlled for in CORA (e.g., prior domain knowledge, attitude toward the task topic). This aspect is crucial since students may pass different credibility judgments and follow diverse reasoning approaches depending on their initial stance (Kahneman et al., 1982; Flanagin et al., 2018). At a later, longitudinal research stage (e.g., in the context of formative assessments; Section Analyses of Response Processes and Longitudinal Studies), attitude-dependent tasks can be administered to assess COR levels among students for topics they explicitly support or oppose. Not solving the task in a way that accounts for both perspectives would therefore yield a lower CORA test score (Section Scoring Rubrics). This in turn would strengthen the high ecological validity of CORA.

Whichever stance the students choose, they will not be awarded points unless they provide warrant through reasons and arguments, and back them up with evidence from the evaluated website and further consulted online sources (Section Scoring Rubrics). Thus, an evaluation supported by reason and evidence (such as a link to an authoritative website), judged by raters as acceptable against a generic or domain-specific quality standard,

is used to infer the extent to which students' have critically reasoned with and about online information. The call to justify is explicitly prompted in the task ("provide a justification") and the backing with evidence is required in a separate field, asking for the URLs of consulted further websites. Providing citations is a common form of evidence in academic writing, and copying a URL does not require an elaborate evidential standard. The reasons and arguments students cite in their written responses are scored for plausibility and validity based on a few rules (e.g., "trusting only the source's own claims about itself is not sufficient reason"). The indicated URLs are also evaluated in terms of their trustworthiness (Nagel et al., 2020). We assume that students with advanced COR abilities cite only the best sources they found and used to back up their argument. Conversely, indicating many relevant and trustworthy sources as well as irrelevant and untrustworthy ones was considered an indicator of reasoning that was not fully sufficient (see scoring rubrics in Section Scoring Rubrics).

According to the fundamentals of argumentation, the main claim, reasons, and backing (e.g., evidence) are the basic elements of a reasonable argument (Toulmin, 2003; Walton, 2006). Hence, we considered indications of these, which are also explicitly prompted in the CORA task, in a somewhat aligned manner in the students' responses as evidence that students performed argumentative reasoning. Some argumentation frameworks include further basic components, such as rebuttal and undercut as types of opposing reasons or inclusion of consequences (Toulmin, 2003). These components can be included in further CORA tasks (Section Refining and Expanding CORA), but were not required for the short online evaluation tasks. Moreover, in terms of metacognitive evaluation, students are expected to engage the evaluative critical reflection, i.e., "self-reflective review" of their task solution after formulating their response to the task.

## Scoring Rubrics

According to the task and the student models, CORA tasks measure whether students employ critical evaluation of trustworthiness and critically reason based on arguments from the online information they used. Based on our prior research on performance assessments of learning (e.g., for the international Performance Assessment of Learning (iPAL) project, see Shavelson et al., 2019) and the developed scoring approach, we created and applied new scoring rubrics focusing on the main facets of COR and on fine-grained differentiations of scoring subcategories in accordance with our construct definition (Section Construct Definition of Critical Online Reasoning; for an excerpt of the facet "weighting arguments," see Table 3).

Each task is scored with a maximum of 2 points, with to 0.5 points awarded if the response mentions a major bias or credibility cue, for instance noticing a (covert) advertising purpose, *and* if its implications for the interpretation of information are identified. Up to 0.5 points are awarded if the students support their claim (no matter which stance) with one or two valid reasons that are weighted in relation to each other, and a maximum of 0.5 points if students refer to one or two credible external sources (that are aligned with their overall

**TABLE 3** | Excerpt of the COR scoring scheme; REAS facet, sub-facet “weighing reasons.”

COR facet	Sub-facet	Description	Subscore “weighing reasons” <sup>a</sup>	1	2	3	4	5
				Not fulfilled	Mostly not fulfilled	Partially fulfilled	Mostly fulfilled	Completely fulfilled
Reasoning using Evidence, Argumentation, and Synthesis	Weighing reasons	<ul style="list-style-type: none"> <li>Balanced judgement: pros and cons/diverse perspectives</li> <li>Conclusive argument, comprehensible justification</li> </ul>	0.5 points total 0.1 points per gradation	0.1 tasks not fulfilled at all	0.2	0.3	0.4	0.5 tasks completely fulfilled

Note: completely fulfilled (full score), mostly (not)/partially fulfilled (few, half or most, but not all aspects covered), not fulfilled (zero points).

<sup>a</sup>The complete scoring scheme describes all major facets with scored sub-facets. This excerpt shows the REAS facet with the sub-facet “weighing reasons”.

argumentation). Furthermore, students can achieve 0.5 points if their response is coherent, clearly related to the task prompt, and covers all sub-parts.

In contrast to a simple trustworthiness judgment, which could be performed without further reflection using heuristics, the underlying analytical reasoning requirements of the tasks are more demanding. It is also possible for participants to take the evidence for their criticism of a website from the website itself as long as the argument is warranted and conclusive. Consequently, the scoring rubrics also consider to what extent the students recognize the specific characteristics for or against the trustworthiness of certain websites, cues, and strands of evidence, and whether they consider them in their reasoning and decision-making processes. A student may identify manipulative techniques “X” and “Y” being used by the linked website, which make it untrustworthy, and cite them from the website. In this case, students can receive points for correctly judging the website as unreliable and for identifying a bias, even if they have not accessed external websites. In a follow-up study, in addition to this holistic score per task, further sub-scores can be awarded at different levels of granularity in accordance with the COR construct definition (Section Development of Scoring Modular Rubrics).

Regarding *information trust* strands of evidence, before scoring this aspect of students' responses, we evaluate the stimuli in the CORA tasks in terms of type, number, and location of cues for/against the credibility of a website (Section Test Definition and Operationalization of COR: Design and Characteristics of CORA). In addition to evaluating the stimuli individually, we mark their valence and importance for main argumentative claims (e.g., supporting or contradicting the trustworthiness of the linked website). Given the large number of possible cues, we make some systematic limitations: the collection of cues is mainly restricted to the stimulus materials to be evaluated by all participants. These cues are listed and scored depending on how frequently they are mentioned in the students' argumentative responses (i.e., focus on cues that students selected). In terms of possible verification of plausibility of reasons, we distinguish first-order reasons (e.g., “the website has an imprint”), which may lead to a successful judgment in certain cases and guard against some deceptions if only cues regarding credibility are used, to second-order reasons (e.g., “any website can have an

imprint nowadays, but the indicated organization cannot be found online”).

Further, the cues were systematized following three strands of evidence in accordance with the 3/S Information Trust model (Lucassen and Schraagen, 2011) and Prominence Interpretation Theory (Tseng and Fogg, 1999), including *surface/design*, *semantics/content*, and *source/operator*. Each of these strands can make a specific contribution to an argument about whether to trust information or not. Moreover, they address different reasoning approaches from “aesthetic” appraisal and consideration of mediated presentation, to content and argumentative appraisal as well as to consideration of authorship reputation, intent and expertise (and other cues of the production/publication process).

In addition to the described strands of evidence, the model was expanded by distinguishing a primary- and a secondary sources perspective for each strand. Usually, both perspectives will be used to some extent for an evaluation of trustworthiness, i.e., when verifying a cue oneself, evidence standards (standards related to the information itself) are used than when relying on other persons' judgments (here, one rather uses standards related to the probability of successful judgment of the other person). For example, when judging trustworthiness of the author, a student may complete their own research on relevant aspects from a variety of biographic sources or they may follow a journalist's understanding of this author. Verifying every aspect oneself marks a fully autonomous learner, though we acknowledge that this may not be feasible for all aspects in the short test-taking time. For each task, strands containing important cues were listed. Moreover, major distractors supporting a competing assumption were marked.

The rating was carried out by at least two trained scorers per task. For the overall CORA test score, i.e., the average scores of two or three raters for each participant and for each task, a sufficient interrater agreement was determined, with Cohen's kappa >0.80 ( $p = 0.000$ ).

## PRELIMINARY VALIDATION

The validation of the CORA was integrated with the ECD and follows the AERA Standards (Section Research Objectives

and Questions). Starting from the holistic nature of the CORA (see section Task Model), the construct specification, and the modular extensions of the scoring in this paper (see *Interpretative Model*), we present preliminary validity evidence related to the content of the construct. After the COR construct specification and the assessment design, the newly developed CORA tasks underwent content analyses and were submitted to expert evaluation during interviews. The aims were to examine the coverage of the theoretically derived COR construct facets by the holistic tasks and to obtain expert judgment regarding the suitability of the content and requirements for higher education in Germany. Below, we outline the methodology (Sections Content Analysis: CORA Task Components as Coverage of the Construct and Expert Interviews) and discuss the results for both analyses (Section Findings From the Expert Interviews and Content Analysis).

## Content Analysis: CORA Task Components as Coverage of the Construct

A qualitative content analysis (Mayring, 2014) of the CORA tasks was carried out by the CORA research team members who participated in the construct specification but not the selection of task stimuli. Task prompts and the encompassed stimuli were examined to determine the presence or absence of features that would allow test-takers to draw inferences and generate responses worth partial or full credit according to the scoring rubric (Section Scoring Rubrics). The six higher education CORA tasks that resulted from the design process (Section Test Definition and Operationalization of COR: Design and Characteristics of CORA) were coded according the following features and underlying (theoretical) frameworks:

- (1) As part of the meta-cognitive facet, activation of COR was coded to gather evidence on whether the tasks tap students' overall COR ability, i.e., whether they convey a need for critical evaluation and argumentative reasoning and at which point: at the beginning, middle or end of the task. We coded for activation by prompt or by context, by specific cues that would highlight the need for COR during task processing, and for end-of-task activation by required (metacognitive) review steps or invited by a contradictory or uncertain preliminary conclusion. The expectation was that at least some tasks would have a cue for COR activation at the beginning of the task, whereas others may only have a mid-task activation to tap students' ability to identify situations when it is needed to activate their COR.

Moreover, the aspect of problem definition (in the sense of the IPS-I model, **Table 2**) was examined. We coded whether the task was embedded in a broader activity context to support judgment based on purpose and increase ecological validity (e.g., judging information trustworthiness *for use in a term paper*); in a pretest during task design, students had claimed to apply more or less rigorous evidence standards depending on purpose. We also coded whether the task goal was clearly stated in the prompt and whether solution criteria were given or if they needed to be inferred.

Other MCA subfacets regarding regulation, affective response, or attitudinal aspects were not coded due to the difficulty of assigning them to specific task features (in the online assessment); these could be elicited more efficiently in a future coglabs study (Section Analyses of Response Processes and Longitudinal Studies).

- (2) The OIA and CIE facets were assumed to be organized in order of the phases of the IPS-I process model to highlight similarities and differences among the CORA tasks, while specific features were coded based on other additional models and research foci (Section Scoring Rubrics). The phases of source selection and initial scanning of a website were listed under one facet (OIA or CIE), but are expected to be hybrid search and evaluation activities (to be further examined in coglabs; Section Analyses of Response Processes and Longitudinal Studies).

Among the search-related aspects (OIA), we coded the necessity to use different search interfaces during the process (e.g., a search engine, in-site search) to obtain reliable information. We assumed that basic search skills, but not use of advanced search operators or special databases would be required. Websites that were inaccessible and media that would not play or were too long and not searchable were excluded during the pretest. Hence, suitable information was expected to be fairly easy to locate and access (except on specific search tasks) by performing an external search.

Regarding information source selection, we generally coded the sources students had to evaluate to obtain suitable information, i.e., the given website, additional websites, and linked sources (e.g., a background article to a tweet), and/or websites which students selected themselves. We expected requirements to vary across tasks.

- (3) The facet of CIE united the IPS-I phases of scanning a website and in-depth information processing. For global website appraisal and orientation, we coded to what extent it was necessary to judge the overall layout and design (or if one could ignore the context and start reading/searching immediately), to what extent students needed to get an overview first, for instance to find a suitable paragraph in time by scanning sub-headings, and determine if they had to attend to any specific cues rather than reading the main text. We expected that some websites might have obvious design cues and others might not (e.g., a popular social network could be interpreted as an obvious cue for lower credibility); some websites were expected to be more complex or longer and require initial orientation; however, we expected students to find relevant information on the given landing page and standard sub-pages (e.g., publisher and author listed in the legal notice or "about" page); we expected the task solutions to not be based solely on the identification of a single cue.

Regarding information processing, we generally assumed the required reading comprehension to be a given among higher education students and focused on evidence evaluation, classifying available cues based on the 3'S' Information Trust model (Lucassen and Schraagen, 2011) into cues in the design, content, or source, as well as (jointly

for all three) secondary external sources indicating cue evaluations (Deferring judgment to external sources would also require an evaluation of these sources' expertise and intent). For example, if a website had aggressive popup advertisement, this would be coded as a cue in the design that might indicate lower trustworthiness. We expected that not all tasks would have cues for (un)trustworthiness in all strands, but at least in one strand of evidence. Moreover, different strands of evidence would be tapped across tasks so that no single subset of evaluation skills or strategy (e.g., only using logical critique or looking up the author's reputation) would be universally successful.

- (4) Based on major components of reasoning (Walton, 2006) with evidence, argumentation and synthesizing (REAS), we coded to what extent students needed to cite sources of evidence (expected), to what extent they had to provide reasons why they trusted the information (on some tasks), or arguments against its trustworthiness (expected), to what extent they needed to make an overall evaluative judgment (expected), and to what extent they had to synthesize and weight possibly contradicting information and arguments (expected), to what extent stimulus materials contained a prominent bias, mismatched heuristic, or fallacy to be avoided (expected for most tasks), and if there was a clear-cut solution vs. an undecidable outcome so they had to account for uncertainty (only on few tasks).

In regard to presentation of results (another IPS-I phase), we coded to what extent the quality of the structure and phrasing of students' responses contributed to their score. As we focused on the quality of argumentative links and information nodes rather than their rhetorical arrangement, we expected response structures and phrasing to not matter beyond the general effort of presenting a coherent and conclusively argued response.

- (5) In addition, given that domain- and topic-dependent prior knowledge (and attitudes) might influence participants' searches, evaluation, and reasoning, we collected some descriptive information on the task topics: We labeled the origin of the misinformation as an indicator of how widespread and hard to identify a deception might be (e.g., from a single author's error on a page to a newspaper editorial board's agenda-setting policy to a culturally normalized conviction), as suggested by the Hierarchy of Influences model (Shoemaker and Reese, 2014). We coded the share of supporting and opposing (in terms of the task solution: conducive or distracting) search results for the key terms in the prompt and website title (as an indicator of controversy and how easy it was to find additional online information). We labeled the broader task context in terms of societal sphere (commerce, science, history etc.), the kind of misinformation genre, specific biases, heuristics, and fallacies presented, and the type of online medium. The overall expectation was that CORA tasks would present one or two challenging aspects but not be overly difficult given the short testing time (e.g., no national scandal to be uncovered), and would be varied in their genre and contexts. Results are summarized in **Table 4**.

## Expert Interviews

Semi-structured expert interviews (Schnell et al., 2011) provided a second source of evidence on content representativeness. In semi-structured interviews with experts, we presented examples of CORA tasks and asked experts to comment *on their suitability for higher education in Germany*. The interviewed experts were leading academics in their field and included two of the U.S. developers of the civic online reasoning assessment, four experts in computer-based performance assessments in higher education, and six scholars from the fields of media studies (who focus on online source evaluation or media literacy), linguistics, and cultural studies. After considering the task stimuli, prompt, and rubrics (sent to them in advance), the experts were given the opportunity to ask for clarifications and were then asked to share their first impressions of the assessment before responding to more specific questions regarding the tasks and features. The discussed topics are shown in **Table 5**.

The questions were asked in view of the German context and tasks specifically, since the media landscape and typical challenges with online information, including deception strategies, can be country-specific. Experts' responses were interpreted in light of their disciplinary backgrounds and convergence or divergence between experts.

## Findings From the Expert Interviews and Content Analysis

In the following, we present a summary of the main findings from the expert interviews and content analysis.

### Overall Experts Evaluations

Overall, with regard to the suitability and validity of the CORA tasks for higher education in Germany, most experts agreed and confirmed the content and ecological validity of the CORA tasks and recommended further expansions. For instance: "The task is clear, the instruction is also clear, and it seems obvious that they need to formulate a response."

One expert, after pondering how to translate and adapt the U.S. tasks, and worrying about cultural suitability, considered the CORA tasks and commented: "These [German] tasks are really a hundred times better for Germany."

### Coverage of COR Facets

One question critically discussed with experts addresses the domain-specificity of the CORA tasks. Here, the experts confirmed that the six tasks cover generic COR ability. For instance: "No domain-specific knowledge is required. It's a good selection for the news/science context."

One concern that was raised by most experts regards the suitability of the testing time to assess all facets of COR, and in particular the REAS facet. However, experts also agreed that students may not dedicate more time to the task when evaluating an information source in a real setting. As one expert notes: "There are 10 min to conduct a search. One may doubt if people would commit as much time in everyday life, unless they really took the time to carry out a more detailed search." At the same time, the natural online environment of the assessment was praised in terms of the high ecological validity of CORA by all

**TABLE 4 |** Content analysis of the CORA tasks as coverage of the major facets of the construct.

COR facet	1	2	3	4	5
<b>A. Metacognitive Activation (MCA)</b>					
<i>COR activation</i>					
Initial activation by prompt	++	++	++	++	++
Initial activation by context	0	0 (+)	2nd	0 (+)	0
Mid-task activation by cue identification	+	opt	2nd	+	2nd
End-task activation by synthesis outcome	0	opt	0	opt	opt
End-task activation by review process	0	0	0	0	0
<i>Problem definition</i>	see IPS-I (Brand-Gruwel et al., 2009)				
Clear purpose of activity	0	0	0	0	0
Clear task goal	++	(+)	++	++	(+)
Determining criteria	+	+	+	+	+
<b>1. Online Information Acquisition (OIA)</b>					
<i>Search</i>	see IPS-I, Interactive Information Retrieval (Xie, 2008)				
Search engine use	opt	+	+	opt	+
Data base use	0	opt	0	0	0
Defining query terms	+	+	+	+	+
Spec. search (e.g., operators)	0	0	0	0	0
In-site search	0	+	0	0	+
<i>Information source selection</i>					
Evaluating given source	+	0	+	+	+
Evaluating linked sources	0	0	opt	0	opt
Evaluating self-selected sources	opt	+	+	+	+
<b>2. Critical Information Evaluation (CIE)</b>					
<i>Scanning (global site appraisal)</i>	see IPS-I				
Of design	+	0	+	+	+
Of structure	+	0	+	0	+
Of spec. features	+	0	opt	0	0
<i>Information processing; evaluation of strands of evidence</i>	see IPS-I; Information Trust (Lucassen and Schraagen, 2011); Prominence Interpretation Theory (Tseng and Fogg, 1999)				
Cues in design	++	0	0	+	0
Cues in content	0	+	+	+	+
Cues on author/publisher	+	0	+	0	+
Cues reported on in external sources (alternatively: expertise and intent)	Opt	++	+	+/-	+
<b>3. Reasoning using Evidence, Argumentation, and Synthesis (REAS)</b> see Fundamentals of Critical Argumentation (Walton, 2006)					
Citing external sources of evidence	opt	+	opt	opt	+
Generating supporting reasons	opt	opt	opt	+	+
Generating opposing reasons	+	+	+	+	+
Making a holistic evaluative judgment	+	+	+	0 (scaff)	+
By synthesizing and weighing information	0	+	+	0	+

(Continued)

**TABLE 4 |** Continued

COR facet	1	2	3	4	5
Avoiding biases, heuristics, fallacies	+	+	+	+	+
Adjusting for uncertainty	0	0	opt	opt	+
Structuring and presenting results			see IPS-I		
Organizing structure	0 (scaff)	0 (scaff)	0 (scaff)	0 (scaff)	0 (scaff)
Formulating response	+	+	+	+	+
<b>B. Descriptive Features Hierarchy of Influences (Shoemaker and Reese, 2014)</b>					
Origin and diffusion of misinformation	organization	personal	professional association	news editor	think tank
Share of supporting/distracting external sources	+/0	+/- -	+/-	0/-	+/-
Social context	DE: commerce (1), politics (2), society/ethics (2), history (0), science (0),... US: commerce (0), politics (3), society/ethics (3), history (1), science (0),...				
Misinformation genres	corporate educational texts, hidden advocacy, tendentious media commentary, social media rumor,...				
Specific biases, heuristics, fallacies	commercial bias (limited information selection, overstatement of pros, understatement of cons), oversimplification of opposing stance, ideological bias (religion, economic policy), unsupported one-sided prediction, baseless ridicule,...				
Media type	news article, website, tweet, Facebook post, and news video				

Features per individual task; descriptive features aggregated across tasks for context, genre, biases, media.

++ = cue is obvious from task (e.g., explicit prompt).

+ = cue is present but needs to be identified and used in inference; i.e., ability likely tapped.

opt = cue is available for use, but not required for optimal solution; i.e., supporting sub-facet likely tapped.

0 = cue not available or offers no information; i.e., applying ability on problem yields no result.

0 (scaff) = requirement scaffolded; i.e., partial solution is given.

- = cues misleading; i.e., distractors present, critical evaluation (selection) likely tapped.

-- = most cues misleading; i.e., distractors present, critical evaluation (selection) and search skills likely tapped.

2nd = cue present, but can be identified only after successful inference, e.g., after deception or bias detection, i.e., critical evaluation and argumentation (weighing of alternative explanations) likely tapped.

**TABLE 5 |** Evaluation questions for experts (selection).

Suitability for higher education in Germany	To what extent are the COR tasks suited to assess students' ability to critically use online information and reason based on it?
Coverage of facets	Which (sub)facets of COR might not be assessed by the tasks?
Representativeness of media	Are the media in the stimuli representative of the information environment higher education students typically encounter online?
Representativeness of misinformation types	Are the kinds of biased information representative of the types/genres of misinformation and biases that students should be able to recognize? Which ones might be missing?
Difficulty and source use	How do you judge the difficulty of tasks?/What university level are they suited for? What aspects might contribute to the difficulty? Do students need to evaluate other sources that are significantly more or less difficult to evaluate?
Potential for differential item functioning (DIF)	Would you expect any group of students to perform better on the tasks (e.g., depending on gender, age, study domain)?/Does the assessment disadvantage any group? Does the assessment have a potential bias?
Item design	Which aspects are particularly important to select or construct realistic tasks with adequate difficulty (to tap COR abilities)?

experts: "The mode of administration as given here is important, since it enables assessing internet search behavior."

The new rating scheme with the subscores and evaluation categories was positively evaluated by the experts, although they stressed the high complexity of the scoring rubrics. For instance: "It is also good that you have different degrees, not only "right or wrong." Of course, this places high demands on coders, but with training, it is doable."

### Representativeness of Media

Most experts positively evaluated the representativeness of the chosen media, i.e., media that students frequently use online. However, one expert criticized that "scientific and journalistic media were indeed covered, but the selection could include more reputable media as well as some media more on the lower quality end of the spectrum. The ones here are well chosen; one cannot immediately tell if they are fabricated or not." Another expert

proposed: “These are common media sources. However, you may include even more social media, and not only evaluate news by institutions and organizations, but also by individual users or from the “alternative” news outlets. Influencers on Instagram who present products are another option.”

### Representativeness of Misinformation Types

In terms of the presented misinformation, the overall judgment by the experts was positive. For instance: “Item topics are nicely varied; tasks are not too simple, so one does not get bored; and I could not decide right away, I had to click on the [background source] and take a look. Even as a media-competent person I had to examine it to make a judgment.” Another expert stated: “I could not solve the items without checking. I had heard nothing about these cases. With unknown issues, ideology also plays a smaller role.” In terms of potential biases and DIFs, the experts did not express any concerns. For instance: “I do not think that, given equal competence, it would be easier for students with typically liberal or left-wing attitudes to solve the tasks. The selection of topics in the tasks covers some stances typically accepted in the left and green camp, some typically accepted by the conservative camp.... It is a good mix.” In addition, one expert recommended expanding the item pool by a clearly untrustworthy website and one clearly trustworthy website, so that lack of trustworthiness would not be predictable on post-tests. Another expert proposed: “Some other frequently shared information of low trustworthiness can include memes, misattributed or completely wrong quotes, or quotes taken out of context.”

### Difficulty and Source Use

At the same time, however, it was questioned whether the task prompts might be too difficult for beginning or undergraduate students. For instance, “Even as a frequent evaluator, I was not always skeptical of the given information.” In this context, the appropriateness of the limited testing time was once again questioned. Only one expert was of the opinion that the tapped skills are mastered early on during the course of studies: “What you assess here is what we call study of sources. [...] We teach this the first year in our degree course, and from then on, students should know it, and it is basically part of practice from then on.” In this regard, some experts recommend splitting the task into parts that focus on particular facets of COR. For instance, “You could ask for an ad hoc judgment, and have additional tasks [for more detailed search].” Another expert proposed: “If students do not find suitable sources, they may get stuck. Perhaps, it is worth including a separate task format or hints.”

Another aspect addressed by most experts concerns participants' prior knowledge, beliefs and critical stances, which may significantly influence their CORA test performance. In this context, one expert stated: “Whether people evaluate sources can also depend on their motivation to put in the time for checking them. Hence, need for cognition could be an influence, people's proclivity to get to the bottom of things and not avoid complexity.” Similarly, another expert commented: “People may also carry out a detailed search just

to confirm their worldview or to form an opinion. This can occur despite existing search skills (but they would still be ideologically stuck). Hence, motivation to be open to other positions is key, and it then matters how much time I'm willing to invest in a search.” In this context, most experts stressed the need to control for participants' prior knowledge and attitudes. For instance: “Political orientation can be used as a control variable if completely anonymized, for instance asking where they would position themselves on a 1-to-10 left-to-right-wing scale (on a voluntary basis) appears less invasive.” Another expert proposed: “You may also want to specify whether it is the successful judgment of a first impression or openness to changing one's opinion. In that case, personality traits would be controlled for. So, a different option would be to assess who changes their mind when they come across new leads.”

### Suggestions for the Further Development of CORA

The interviewees did not recommend the exclusion of any tasks. In few cases, the experts recommended removing certain task features. However, the experts provided a number of recommendations in terms of refining the CORA. For instance: “The role of content shared by friends could be expanded, where it is unclear if it has been checked or not... User comments can be read and might influence more passive users... So to increase difficulty, you could add social credibility cues. It would be an even more realistic setting, but you need to see how additional information would influence difficulty.” This suggestion is in line with credibility research that highlights the huge role that social persuasion by peers plays in today's social media (Fogg, 2003). Although cues exist in few of the tasks, social persuasion and learning was purposely left for future CORA expansions (Section Refining and Expanding CORA).

### Overview Content Analysis

As task prompts shared a similar structure and wording with only differing topics and source links, the evaluative and argumentative requirements were assumed to be similar as well. The closer content analysis, however, revealed two distinct types of tasks: (1) “website evaluation tasks,” tapping particularly CIE, but less OIA if students did not search beyond the presented website; and (2) “fact-checking tasks” that only presented a claim, but no linked website as a stimulus, and therefore forced an Internet search. Fact-checking emphasized OIA more in comparison to CIE since students were not bound to evaluate one particular website; if they were uncertain about a source, they could abandon it to find a better alternative. In this way, the task types afforded use of all three facets but, respectively prioritized one in particular; consequently, a third format emphasizing REAS to complement the other two would be a further development step.

The task response sheet provides students with a clear structure, with the sections overall trustworthiness judgment, warrant (sometimes with separate pro and con sections), and URLs. The scoring rubrics did not contain any specific language requirements. Nonetheless, the students had to fill in the response

sheet sections in a coherent way and formulate a conclusive statement to be awarded points. While the strands of evidence varied systematically, content-related aspects and difficulty were not systematically varied across tasks. At the current stage, given the large number of available topics and types of biases, it is still a small, to-be-expanded task pool (Section Refining and Expanding CORA).

Regarding the individual COR facets, the content analysis showed the following findings:

### Metacognitive Activation (MCA)

In terms of the activation of COR, all task prompts offered clear instructions to evaluate the trustworthiness of the sources at the beginning of the task. Mid-task activation depended on the presence of specific cues. All tasks contained at least one explicit initial and one implicit mid-task cue that might alert students to the need to use their COR. End-of-task activation, for instance a prompt to explicitly review and reflect, was not employed. Moreover, there were no tasks with implicit mid-task or end-of-task activation only, which is a characteristic of deception in online information in real life (i.e., there rarely are prior warnings that a website might contain *misinformation*, compared to automated warnings and filters for, e.g., malware detection). The primary aim of CORA is to measure performance during Internet searches, critical evaluation, and argumentative reasoning; it would be hardly possible to assess these facets if students missed the activation cue.

In regard to the aspect of “problem definition,” while problems were clearly stated in the task prompts, the students need to determine the evaluation criteria for trustworthiness and untrustworthiness themselves. Some experts critically noted that students may be unsure about the required evidence standards. It remains an open question whether deriving criteria for one's trustworthiness judgment should be part of the COR ability. This aspect has been scaffolded in some think-aloud studies, though we are not aware of scaffolding in other assessments. In think-aloud studies, the evaluation of criteria has been separated into different steps; for instance consecutive filtering of sources based first on relevance, then trustworthiness, then usefulness (Walraven et al., 2009; Goldman et al., 2013).

### Online Information Acquisition (OIA)

Regarding expected search skills, content analyses indicated that students can find a suitable source, and in one task a complete website review, even without specific search terms apart from the titles as long as they searched for external sources at all. Only for the fact-checking task did we find an expected larger share of relevant distracting SERP results. For selection of sources for reading, features also varied as expected. Even though some stimuli are quite short (e.g., a tweet), not all students may open the linked background article with more information. However, as this is clearly included as the main piece of evidence backing up the claim in the stimulus, students' attention to the link as a cue and to the background article can be considered a legitimate part of the tapped COR ability. An examination of the SERPs for major keywords showed significant variation in terms of the available information on the first SERP page and across tasks

(see section Descriptive Features), which usually included some supporting, but also multiple irrelevant or misleading sources on the first page. Thus, the tasks appear to tap students' skills in SERP evaluation, as desired; for instance, students need to actively decide which websites to focus on.

### Critical Information Evaluation (CIE)

As expected, some websites contain too much text to process in the limited time and require students to search for or skim the content. Most webpages contained more text on the landing page than fit on a single screen, and had common sub-pages, such as the “legal notice” or “about” section. For their own orientation and for a fast trustworthiness judgment, students need to gain a comprehensive overview of the websites first to be able to deliberately focus on specific sections. Some tasks also required students to recognize and understand cues outside the main text (e.g., an organization logo at the top). This indicates that simply starting to read the text might be an unsuccessful strategy on these tasks and would take too much time.

In terms of strands of evidence, cues were well distributed across tasks, in fact more regularly than expected. There were usually at least two strands of evidence with relevant cues available, so students could take different routes through the task. The linked background webpages usually contained cues that need to be understood and evaluated. Suitable information was also available in (purposefully selected) external sources to help students solve the tasks and, for instance verify the reputation of an unknown author. While tasks could be solved using just one of the available strands of evidence (e.g., only cues on author), combining two or more converging strands could potentially afford higher confidence in task response and possibly minimize effects of interpretation errors. This supports the intended interpretation of task performance.

### Reasoning Based on Evidence, Argumentation, and Synthesis (REAS)

In terms of the argumentative component of COR, students needed to make a judgment in all tasks, mostly by weighting the pros and cons, although some tasks also scaffolded these, asking for both pros and cons separately rather than a final integrated decision. These requirements can be varied more systematically based on empirical evidence regarding task difficulty. All tasks required students to find *disconfirming* evidence or arguments, which supports the interpretation that “critical” reasoning skills are tapped, and some tasks required students to find both confirming and disconfirming evidence or arguments. This could place students who rely only on their confirmation bias at a disadvantage, as intended. However, one expert called for the inclusion of clearly trustworthy or untrustworthy websites to better discriminate performance at the lower skill range and prevent re-testing effects (i.e., students assuming that all websites in the assessment are untrustworthy). While citing external sources is required on all tasks and is often beneficial to building an evidence-based argument, it implies a certain trade-off, as evaluating these external sources takes time and requires a higher cognitive effort. A REAS-focused task format might juxtapose several pre-selected sources with potentially

contradictory information that would need to be argumentatively weighted and synthesized. Such tasks have been developed in the iPAL project (Shavelson et al., 2019; Zlatkin-Troitschanskaia et al., 2019) and MSC research (for an overview, see Braasch et al., 2018).

### Descriptive Features

Task topics were varied, as expected, albeit not all societal spheres were equally covered—experts did not judge this aspect as particularly important. Sources of misinformation, including websites by associations and individuals, small enterprises, and editorial teams, were at the lower to medium level in the Hierarchy of Influences. These sources of misinformation were still mostly identifiable as individual entities within a pluralistic information environment. The CORA tasks did not focus on entities at higher levels of influence, such as media corporations or government agencies. Thus, there were no high-level “scandals” involved (as often referenced in conspiracy-related misinformation). This may imply that the highest levels of COR related to analyzing societal and funding contexts, as typically required from investigative journalists, are not tapped in the CORA tasks. This is reasonable given the time limit and the lack of content-related knowledge requirements. However, as the experts noted, the selection of topics, contexts, and genres covered in the tasks could be varied more systematically (Section Refining and Expanding CORA).

### Summary

The preliminary validity evidence from content analysis and expert interviews indicated some important implications for the CORA. Overall, both the content analysis and the expert interviews indicated that it taps higher education students' abilities to search, access, critically evaluate, use, and reason based on online information in the German context, with a slightly stronger focus on the evaluation components. The preliminary evidence supports our validity claim that the CORA measures the participants' personal construct-relevant abilities in the sense of the defined construct definition (RQ3). Moreover, the expert interviews indicated that the CORA tasks cover a significant portion of the online media landscape relevant for higher education students in Germany as well as typical problems and genres of websites and online texts that call for COR skills in the German higher education context.

## RESEARCH PERSPECTIVES

### Refining and Expanding CORA

The CORA tasks allow for a variety of more detailed (sub)scores, for instance as adaptive feedback based on the navigation logs. Two crucial dimensions that are underrepresented in the scoring so far are the metacognitive activation of the COR abilities in relevant contexts and situations as well as the reviewing of own task-related knowledge and beliefs. This important aspect of COR also aligns with the activation of epistemic metacognition, and can offset own cognitive heuristics (e.g., confirmation bias). COR activation is currently triggered by the task prompt and tapped by the CORA tasks. Sub-tasks

could be developed to assess COR activation in a more focused manner, for instance by using a format that assesses context-dependent choice of action (e.g., decision to evaluate a website or not) as tapped by situational judgment tests (Weekley and Ployhart, 2013). There is also potential for task prompts to include an explicit purpose of the activity indicating subsequent use of the evaluated information. However, as Goldman and Brand-Gruwel (2018) stress, future research might need to focus more intensively on psychological stimuli embedded in the tasks and their complex interrelation with the task solvers' response processes that these stimuli might activate (e.g., rereading, thinking critically).

Metacognitive reviewing of (prior) knowledge and beliefs, accepts the high probability that students do not withstand every manipulation attempt and have likely already acquired prior knowledge based on misinformation, which can only be transformed if it is reconsidered in light of the new knowledge. If an inconsistency between new (warranted) knowledge and prior (misinformed) knowledge occurs, it can only be resolved in an epistemically justified way if the prior misinformed knowledge is altered—the opposite might lead to further misconceptions or motivated reasoning. This can be linked to the epistemic virtue of open-mindedness and implicates that negative experiences and failures can provide unique insights for learners and can be transformed into in-depth knowledge in the future (Oser, 2018), but only if reviewed and successfully reinterpreted by the learner. Hence, conducting an open-minded metacognitive review of prior knowledge and beliefs forms a key component of COR, activated by prompts in CORA tasks.

As another direction for further research, in addition to the generic COR assessment, domain-specific CORA tasks have been developed based on the iPAL assessment framework for specific domains (e.g., economics; Zlatkin-Troitschanskaia et al., 2019, 2020a). Since learning environments and the media used by learners within disciplines change with increasing speed due to digitalization and university students' increasing use of information available on the Internet for their domain learning, we will particularly focus on information gathering and knowledge building from mass and social media when further expanding the assessment of domain-specific COR.

### Development of Scoring Modular Rubrics

A sub-score can be awarded per each single/individual aspect in a facet of COR; that is, for activation of COR, the phases during which a trustworthiness judgment is performed (or not), and additional evaluation process (e.g., a fact-checking search) are initiated (or not) (Section Scoring Rubrics). For the critical evaluation facet of COR, scoring can be extended depending on the strands of evidence used, based on the information trust model (Lucassen and Schraagen, 2011, 2013). Collecting evidence from the three strands—(i) on the author, (ii) design/text surface, and (iii) the content—a more reliable evaluation and reasoning than evidence from only one would be awarded a higher score. Similarly consulting external sources and other's judgments of the same aspect would be awarded a higher score than considering only one. The ratio of self-examined vs. externally consulted vs.

not considered strands of evidence can serve as an indicator of (topic-dependent) intellectual autonomy (Paul and Elder, 2005).

Identifying the (possibly hidden) purpose of a website (e.g., sales, political opinion-forming) is a primary phase in the task-solving process. This also includes recognition and understanding of advertisements and other surface features (e.g., authorship). If these behavior- and process-related facets are included in the scoring categories, a time-sequential diagnosis of the quality of online reasoning is possible. These sub-scores can be further used as a basis for the development of adaptive feedback for teachers and students, which indicates when a student is more or less successful in systematically solving a task (or, e.g., they were spending too much time on searching or on one website).

For the argumentation facet of COR, the score can be further differentiated based on use of each argument sub-component: i.e., central claim, reasons, evidence—and implications for task requiring a recommendation (Walton, 2006; Fischer, 2018). A pool of supporting and attacking reasons can be collected from students' responses, weighted in their contribution to task, and used to score subsequent responses (e.g., depending on whether students' used the most weighted reasons, pros and cons, and only claimed an evidence-orientation, or cited evidence, verified evidence, generated own evidence).

A subscore can be awarded on the level of comprehension and reasoning of single text units. This requires a classification of cues at the text surface as an indication of trustworthy or untrustworthy sources and information. At the moment, this can be efficiently analyzed only for websites given as stimuli in the CORA tasks. The quality of additional websites used by the students can only be estimated based on their URLs. Analyzing the quality of all websites the students accessed while solving the CORA tasks would require comprehensive media-specific and content-qualitative analyses as well as in-depth linguistic and computer linguistic analyses (e.g., text mining). In addition, process data, for instance eye-tracking or navigation logs, can be used to support an on-task detection of cues the student has been exposed to (navigation). Similarly, in the REAS facet, single inferences and conclusions presented in the text, indicating author biases, fallacies, and heuristics can be classified and scored depending on whether students' repeat them uncritically in their responses, avoid them, or qualify them. Against the background of familiarity and a critical approach to the topic given in a task, successful students should not copy statements so much as express their own argument and opinion.

Based on prior research identifying different navigation and reasoning profiles (List and Alexander, 2017), respondents could be classified into specific COR "learner profiles" based on their (sub)scores on facets of the rubric (e.g., using cluster analysis). Based on students' initial stance toward a task topic (for, against, neutral), (self-estimated) prior task- and topic-related knowledge (expert, novice), and topic interest (interested or not), students can be distinguished into distinct initial profiles, for instance "novice in favor" or "expert neutral," which may impact students' information search and reasoning approach while solving the CORA tasks: "Novices" may need to form

an initial stance and identify trustworthy references or experts whose judgment they trust, while "experts" may draw on their knowledge of trustworthy sources, or prior reasoning on the topic, but are challenged to not fall for confirmation bias and need to test their position (self-critically) against opposing views. "Novices" may also adopt a naive strategy of no initial evaluation of online information, but fallibilism over time, for instance compensating low evaluation skills with sophisticated epistemic beliefs and thus being open-minded to change their beliefs based on new evidence (Paul and Elder, 2005). Taking into account a longitudinal learning perspective (Section Analyses of Response Processes and Longitudinal Studies), online reasoning can later also include a meta-cognitive facet of less well-known yet important properties that influence students' learning and mental functioning (e.g., built-in gratification mechanisms and resulting media preferences).

Scoring for formative purposes in educational practice can then focus on certain features in students' response processes (Section Analyses of Response Processes and Longitudinal Studies) depending on initial "learner profile" (e.g., presence of pros and cons in responses of "topic experts" vs. "novices"), whereby the profiles may vary depending on the topic tapped by the CORA task. CORA tasks can be retested across several measurement points over a course of students' studies (Section Analyses of Response Processes and Longitudinal Studies). Here, knowledge tasks (e.g., selected-response items) on key pieces of information and misinformation in CORA task stimuli can be used to control for (prior) knowledge or retesting effects, which would be especially important for domain-specific CORA tasks. A pre-post design can indicate domain learning over time, i.e., when a student accepted misinformation on the pretest but did not accept it on the posttest (e.g., indicating misconceptions or conceptual change). The formative assessments can inform teachers and students how they can improve their search and evaluation behavior and domain-learning using the Internet.

## Analyses of Response Processes and Longitudinal Studies

Given the open-information environment and holistic nature of the performance assessment, a number of more detailed analyses of the information environment and students' navigation thereof is being conducted. We aim to connect the assessment design and outcomes to the complex *Information Landscape (IL)* that the individual student encounters online, and examine how it influences the response process and test result (Nagel et al., 2020). Using logged CORA performance data, the students' browsing activity can be examined to describe which sources they accessed, how much time they spent, what judgments they made, and which cues they considered during which phases (Schmidt et al., 2020).

According to the ECD (Mislevy and Haertel, 2006), response processes indicate which cognitions are generated by a confrontation of a subject (student) with a task. The analysis of the response processes can refer to various indicators that arise during the processing of the CORA tasks (e.g., as described

in **Table 2** in Section Test Definition and Operationalization of COR: Design and Characteristics of CORA, with a focus on quality judgments by IPS-I phases). The log files or think-aloud data can give an indication of the expected (meta)cognitive processes that are elicited during the response processes (Zumbo and Hubley, 2017), for instance on the occurrence of different mental processes, students' attention to particular aspects, and their distribution across the task solving phases to determine whether the theoretically assumed (construct-related) comprehension and reasoning processes were indeed performed by respondents.

In a longitudinal analysis perspective, we aim to investigate the relationship between the students' COR ability and their acquisition of reliable warranted vs. erroneous knowledge over the course of their studies in higher education. Using repeated CORA measurements (i.e., formative assessments), aspects of knowledge development and memory (incl. retesting) effects over the course of study can be analyzed, providing an important basis for instructional interventions in educational practice.

## CONCLUSION

The holistic task format allows for modular extensions of sub-scores, provided abilities are tapped, which can be deployed efficiently in subsequent in-depth validation studies. As Goldman and Brand-Gruwel (2018) conclude for sourcing, which equally applies to trustworthiness evaluation and reasoning based on online information more generally: "We also need a more nuanced approach to the purpose and value of sourcing processes; identifying the perspective of a particular source is not the "end goal." Perspective is not so much about trustworthiness of sources as it is about how perspective informs what learners make of the information with respect to forming interpretations, making decisions, and proposing solutions."

We agree and add that, beyond specific text-types dedicated to arguing about trustworthiness (e.g., research papers, legal opinions), trustworthiness evaluation mainly serves to filter out untrustworthy information. That is, hardly any additional information is added that helps students resolve an information problem, and instead available evidence for reasoning that turns out to be untrustworthy is even detracted from an argument. This can appear demotivating to the novices, unless it supports the achievement of a higher-order goal, such as maintaining a high quality standard. In general, learning based on erroneous knowledge may result in either unverified adoption or incorrectly understood or recognized information that can lead to persistent misconceptions and knowledge inconsistencies, which can become evident in later use of this erroneous knowledge.

With the present COR conceptualization and its assessment framework combining information acquisition, trustworthiness evaluation, and argumentative reasoning, we contribute to a better understanding of how trustworthiness judgments are functionally embedded in the broader information acquisition and online reasoning process, and open up perspectives for long-term studies in this emerging research field.

At the same time, this study is only a starting point for longer-term research on critical reasoning at the higher education level within the specific context of the online information environment, which also marks its limitations. Future research would need to determine the relations with critical thinking skills assessed in other contexts as well as with the other cited, partially overlapping assessments (e.g., iPAL performance assessments) that served as a basis and inspiration in the development of the COR assessment.

## DATA AVAILABILITY STATEMENT

The original contributions generated for the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

## ETHICS STATEMENT

Ethical review and approval for the study on human participants was not required in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by participants. Participant statements were anonymized, published in small excerpts only, and checked to not reveal identifiable information.

## AUTHOR CONTRIBUTIONS

DM co-developed the assessment, conducted the analyses, and co-wrote the manuscript. OZ-T co-developed the assessment, supervised the analyses, and co-wrote the manuscript. M-TN co-developed the rating scheme and was involved in preparing, reviewing, and revising the manuscript. SB was involved in preparing, reviewing, and revising the manuscript. SS co-developed the assessment and was involved in its validation. RS was involved in the assessment validation and in reviewing and revising the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was part of the PLATO program, funded by the Rhine-Main Universities fund. Open access publication was supported by the German Research Foundation (DFG) and the Open Access Publication Fund of Humboldt-Universität zu Berlin.

## ACKNOWLEDGMENTS

We would like to thank all members of the Stanford History Education Group as well as all experts who supported this study. We would like to thank the two reviewers and the journal editor who provided helpful critical and constructive comments on the manuscript. We would like to thank the *Frontiers in Education* editorial staff for quality assurance.

## REFERENCES

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., et al. (2008). Instructional interventions affecting critical thinking skills and dispositions: a stage I meta-analysis. *Rev. Educ. Res.* 78, 1102–1134. doi: 10.3102/0034654308326084
- AERA, APA, and NCME. (2014). *Standards of Educational and Psychological Testing*. Washington, DC: AERA, APA, and NCME American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.
- Akamine, S., Kato, Y., Inui, K., and Kurohashi, S. (2008). "Using appearance information for web information credibility analysis," in *2nd International Symposium on Universal Communication, 2008: ISUC 2008; 15–16 December, 2008, Osaka, Japan* (Piscataway, NJ: IEEE), 363–365. doi: 10.1109/ISUC.2008.80
- American Library Association (2000). *Information Literacy Competency Standards for Higher Education*. Available online at: <http://www.ala.org/acrl/standards/informationliteracycompetency> (accessed June 24, 2020). doi: 10.5860/crln.61.3.207
- Arazy, O., and Kopak, R. (2011). On the measurability of information quality. *J. Am. Soc. Inf. Sci. Technol.* 62, 89–99. doi: 10.1002/asi.21447
- Arffman, I. (2007). *The problem of equivalence in translating texts in international reading literacy studies: a text analytic study of three English and Finnish texts used in the PISA 2000 reading test* (dissertation). University of Jyväskylä, Jyväskylä, Finland.
- Banerjee, M., Zlatkin-Troitschanskaia, O., and Roeper, J. (2020). Narratives and their impact on students' information seeking and critical online reasoning in higher education economics and medicine. *Front. Educ.* 5:625. doi: 10.3389/educ.2020.570625
- Batista, J. C. L., and Marques, R. P. F. (2017). *Information and Communication Overload in the Digital Age*. Hershey, PA: IGI Global. doi: 10.4018/978-1-5225-2061-0.ch001
- Bayer, J., Bitiukova, N., Bárd, P., Szakács, J., Alemanno, A., and Uszkiewicz, E. (2019). *Disinformation and Propaganda – Impact on the Functioning of the Rule of Law in the EU and Its Member States. Directorate General for Internal Policies of the Union, Policy Department for Citizens' Rights and Constitutional Affairs*. Available online at: [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL\\_STU\(2019\)608864\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf) (accessed June 24, 2020).
- Beck, K. (2020). "On the relationship between "Education" and "Critical Thinking,"" in *Frontiers and Advances in Positive Learning in the Age of InformaTiOn (PLATO)*, ed O. Zlatkin-Troitschanskaia (Cham: Springer International Publishing), 73–87.
- Blummer, B., and Kenton, J. M. (2015). *Improving Student Information Search: A Metacognitive Approach*. Amsterdam: Elsevier. doi: 10.1533/9781780634623.23
- Braasch, J. L. G., and Bråten, I. (2017). The discrepancy-induced source comprehension (D-ISC) model: basic assumptions and preliminary evidence. *Educ. Psychol.* 52, 167–181. doi: 10.1080/00461520.2017.1323219
- Braasch, J. L. G., Bråten, I., and McCrudden, M. T. (2018). *Handbook of Multiple Source Use*. New York, NY: Routledge Taylor and Francis Group. doi: 10.4324/9781315627496
- Brand-Gruwel, S., Wopereis, I., and Vermetten, Y. (2005). Information problem solving by experts and novices: analysis of a complex cognitive skill. *Comput. Hum. Behav.* 21, 487–508. doi: 10.1016/j.chb.2004.10.005
- Brand-Gruwel, S., Wopereis, I., and Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Comput. Educ.* 53, 1207–17. doi: 10.1016/j.compedu.2009.06.004
- Braten, I., Stadler, M., and Salmeron, L. (2018). "The role of sourcing in discourse comprehension," in *Handbook of Discourse Processes*, eds M. F. Schober, D. N. Rapp, and M. A. Britt (New York, NJ: Taylor and Francis), 141–166. doi: 10.4324/9781315687384-10
- Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., et al. (2019). *Students' Civic Online Reasoning: A National Portrait. Stanford History Education Group and Gibson Consulting*. Available online at: <https://purl.stanford.edu/gf151tb4868> (accessed June 25, 2020).
- Bulger, M. E., Mayer, R. E., and Metzger, M. J. (2014). Knowledge and processes that predict proficiency in digital literacy. *Reading Writing* 27, 1567–1583. doi: 10.1007/s11145-014-9507-2
- Catalano, A. (2013). Patterns of graduate students' information seeking behavior: a meta-synthesis of the literature. *J. Doc.* 69, 243–274. doi: 10.1108/00220411311300066
- Center for Humane Technology (2019). *Ledger of Harms*. Available online at: <https://ledger.humanetech.com/> (accessed October 17, 2019).
- Chen, S., and Chaiken, S. (1999). "The heuristic-systematic model in its broader context," in *Dual-Process Theories in Social Psychology*, eds S. Chaiken and Y. Trope (New York, NY: Guilford Press), 73–96.
- Choi, W. (2015). *A new framework of web credibility assessment and an exploratory study of older adults' information behavior on the web* (dissertation). Florida State University, Tallahassee, FL, United States.
- Ciampaglia, G. L. (2018). "The digital misinformation pipeline," in *Positive Learning in the Age of Information*, eds O. Zlatkin-Troitschanskaia, G. Wittum, and A. Dengel (Wiesbaden: Springer), 413–421. doi: 10.1007/978-3-658-19567-0\_25
- Coiro, J. (2003). Exploring literacy on the internet: reading comprehension on the internet: expanding our understanding of reading comprehension to encompass new literacies. *Reading Teach.* 56, 458–464.
- Damico, J. S., and Panos, A. (2018). Civic media literacy as 21st century source work: future social studies teachers examine web sources about climate change. *J. Soc. Stud. Res.* 42, 345–359. doi: 10.1016/j.jssr.2017.10.001
- Daniels, J. (2009). Cloaked websites: propaganda, cyber-racism and epistemology in the digital era. *N. Media Soc.* 11, 659–683. doi: 10.1177/1461444809105345
- Davey, T., Ferrara, S., Holland, P. W., Shavelson, R., Webb, N. M., and Wise, L. L. (2015). *Psychometric Considerations for the Next Generation of Performance Assessment: Report of the Center for K-12 Assessment and Performance Management at ETS*. Available online at: [https://www.ets.org/Media/Research/pdf/psychometric\\_considerations\\_white\\_paper.pdf](https://www.ets.org/Media/Research/pdf/psychometric_considerations_white_paper.pdf) (accessed July 22, 2018).
- De Neys, W. D. (2006). Dual processing in reasoning: two systems but one reasoner. *Psychol. Sci.* 17, 428–433. doi: 10.1111/j.1467-9280.2006.01723.x
- Dunbar, N. E., Connelly, S., Jensen, M. L., Adame, B. J., Rozzell, B., Griffith, J. A., et al. (2014). Fear appeals, message processing cues, and credibility in the websites of violent, ideological, and nonideological groups. *J. Comput. Mediated Commun.* 19, 871–889. doi: 10.1111/jcc4.12083
- Eisenberg, M. B., and Berkowitz, R. E. (1990). *Information Problem-Solving: The Big Six Skills Approach to Library and Information Skills Instruction*. Norwood, NJ: Ablex.
- Elder, L., and Paul, R. (2010). *Critical Thinking Development: A Stage Theory: With Implications for Instruction*.
- Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educ. Leadersh.* 43, 44–48.
- Evans, J. S. B., and Stanovich, K. E. (2013). Dual-process theories of higher cognition: advancing the debate. *Perspect. Psychol. Sci.* 8, 223–241. doi: 10.1177/1745691612460685
- Facione, P. A. (1990). Critical thinking: a statement of expert consensus for purposes of educational assessment and instruction: executive summary. *The Delphi Report* (accessed June 25, 2020).
- Fischer, F. (2018). *Scientific Reasoning and Argumentation: The Roles of Domain-Specific and Domain-General Knowledge*. New York, NY: Routledge.
- Fischer, F., Chinn, C., Engelmann, K., and Osborne, J. (2018). *Scientific Reasoning and Argumentation: The Roles of Domain-Specific and Domain-General Knowledge, 1st Edn*. London: Routledge. doi: 10.4324/9780203731826-1
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., et al. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Front. Learn. Res.* 2, 28–45. doi: 10.14786/flr.v2i2.96
- Fisher, K. E., Erdelez, S., and McKechnie, L. (2005). *Theories of Information Behavior, ASIST Monograph Series*. Medford, NJ: Information Today.
- Flanagin, A. J., and Metzger, M. J. (2014). "Digital media and perceptions of source credibility in political communication," in *The Oxford Handbook of Political Communication*, eds K. Kenski, and K. Hall (Oxford: Oxford University Press), 417–436. doi: 10.1093/oxfordhb/9780199793471.013.65
- Flanagin, A. J., Metzger, M. J., and Hartsell, E. (2010). *Kids and Credibility: An Empirical Examination of Youth, Digital Media Use, and Information Credibility*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/8778.001.0001

- Flanagin, A. J., Winter, S., and Metzger, M. J. (2018). Making sense of credibility in complex information environments: the role of message sidedness, information source, and thinking styles in credibility evaluation online. *Inf. Commun. Soc.* 23, 1038–1059. doi: 10.1080/1369118X.2018.1547411
- Flore, M., Balahur, A., Podavini, A., and Verile, M. (2019). *Understanding Citizens' Vulnerability to Disinformation and Data-driven Propaganda*. Luxembourg: Publications Office of the European Union. doi: 10.2760/919835
- Fogg, B. J. (2002). *Stanford Guidelines for Web Credibility. A Research Summary From the Stanford Persuasive Technology Lab*. Available online at: www.webcredibility.org/guidelines (accessed June 24, 2020).
- Fogg, B. J. (2003). *Persuasive Technology: Using Computers to Change What We Think and Do*. Amsterdam; Boston: Morgan Kaufmann. Available online at: http://www.loc.gov/catdir/description/els031/2002110617.html (accessed June 24, 2020).
- Fogg, B. J., Marshall, J., Kameda, T., Solomon, J., Rangnekar, A., and Boyd, J. (2001a). "Web credibility research," in *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, ed M. Tremaine (New York, NY: ACM), 295. doi: 10.1145/634067.634242
- Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., et al. (2001b). "What makes web sites credible? A report on a large quantitative study," in *Proceedings of CHI '01: The SIGCHI Conference on Human Factors in Computing Systems*, eds J. Jacko, and A. Sears (New York, NY: ACM Press), 61–68. doi: 10.1145/365024.365037
- Fogg, B. J., Marshall, J., Osipovich, A., Varma, C., Laraki, O., Fang, N., et al. (2000). "Elements that affect web credibility: early results from a self-report study," in *Chi '00 Extended Abstracts on Human Factors in Computing Systems*, ed M. Tremaine (New York, NY: ACM), 287–288. doi: 10.1145/633292.633460
- Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., and Tauber, E. R. (2003). "How do users evaluate the credibility of web sites?," in *Proceedings of the 2003 Conference on Designing for User Experiences*, ed J. Arnowitz (New York, NY: ACM), 1–15. doi: 10.1145/997078.997097
- Gasser, U., Cortesi, S., Malik, M., and Lee, A. (2012). *Youth and Digital Media: From Credibility to Information Quality*. Cambridge, MA: The Berkman Center for Internet and Society. doi: 10.2139/ssrn.2005272
- George, J. F., Giordano, G., and Tilley, P. A. (2016). Website credibility and deceiver credibility: expanding prominence-interpretation theory. *Comput. Hum. Behav.* 54, 83–93. doi: 10.1016/j.chb.2015.07.065
- George, J. F., Tilley, P., and Giordano, G. (2014). Sender credibility and deception detection. *Comput. Hum. Behav.* 35, 1–11. doi: 10.1016/j.chb.2014.02.027
- Go, E., You, K. H., Jung, E., and Shim, H. (2016). Why do we use different types of websites and assign them different levels of credibility? Structural relations among users' motives, types of websites, information credibility, and trust in the press. *Comput. Hum. Behav.* 54, 231–239. doi: 10.1016/j.chb.2015.07.046
- Goldman, S., Lawless, K., Pellegrino, J., Manning, F., Braasch, J., and Gomez, K. (2013). "A technology for assessing multiple source comprehension: an essential skill of the 21st century," in *Technology-Based Assessments for 21st Century Skills: Theoretical and Practical Implications From Modern Research*, eds M. C. Mayrath, J. Clarke-Midura, and D. H. Robinson (Charlotte, NC: Information Age Publishing), 171–207.
- Goldman, S. R., and Brand-Gruwel, S. (2018). "Learning from multiple sources in a digital society," in *International Handbook of the Learning Sciences*, eds F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, and P. Reimann (London: Routledge), 86–95. doi: 10.4324/9781315617572-9
- Goldstein, D. G., and Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychol. Rev.* 109, 75–90. doi: 10.1037/0033-295X.109.1.75
- Gronchi, G., and Giovannelli, F. (2018). Dual process theory of thought and default mode network: a possible neural foundation of fast thinking. *Front. Psychol.* 9:1237. doi: 10.3389/fpsyg.2018.01237
- Hahnel, C., Kroehne, U., Goldhammer, F., Schoor, C., Mahlow, N., and Artelt, C. (2019). Validating process variables of sourcing in an assessment of multiple document comprehension. *Br. J. Educ. Psychol.* 89, 524–537. doi: 10.1111/bjep.12278
- Halpern, D. F. (2014). *Thought and Knowledge: An Introduction to Critical Thinking, 5th Edn.* New York, NY: Psychology Press. doi: 10.4324/9781315885278
- Harkness, J. A. (2003). "Questionnaire translation," in *Cross-Cultural Survey Methods*, eds J. A. Harkness, F. van de Vijver, and P. P. Mohler (Hoboken, NJ: John Wiley and Sons), 35–56.
- Head, A., and Eisenberg, M. B. (2009). Project information literacy progress report: "lessons learned": how college students seek information in the digital age. *SSRN Electron. J.* doi: 10.2139/ssrn.2281478
- Herman, E. S., and Chomsky, N. (2002). *Manufacturing Consent: The Political Economy of the Mass Media*. New York, NY: Pantheon Books.
- Hilligoss, B., and Rieh, S. Y. (2008). Developing a unifying framework of credibility assessment: construct, heuristics, and interaction in context. *Inf. Process. Manag.* 44, 1467–1484. doi: 10.1016/j.ipm.2007.10.001
- International Test Commission (ITC) (2017). *The ITC Guidelines for Translating and Adapting Tests, 2nd Edn.* Available online at: www.intestcom.org (accessed June 24, 2020).
- Jahn, D. (2012). *Kritisches Denken fördern können: Entwicklung eines didaktischen Designs zur Qualifizierung pädagogischer Professionals [Fostering critical thinking: developing a didactic design for qualification of pedagogical professionals]* (dissertation), Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.
- Jahn, D., and Kenner, A. (2018). "Critical thinking in higher education: how to foster it using digital media," in *The Digital Turn in Higher Education*, eds D. Kergel, B. Heidkamp, P. K. Telléus, T. Rachwal, and S. Nowakowski (Wiesbaden: Springer), 81–109. doi: 10.1007/978-3-658-19925-8\_7
- Jozsa, E., Komlodi, A., Ahmad, R., and Heccegfi, K. (2012). "Trust and credibility on the web: the relationship of web experience levels and user judgments," in *IEEE 3rd international conference on cognitive Infocommunications (CogInfoCom)* (Piscataway, NJ: IEEE), 605–610. doi: 10.1109/CogInfoCom.2012.6422051
- Juvina, I., and van Oostendorp, H. (2008). Modeling semantic and structural knowledge in web navigation. *Discourse Process.* 45, 346–364. doi: 10.1080/01638530802145205
- Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511809477
- Kakol, M., Nielek, R., and Wierzbicki, A. (2017). Understanding and predicting web content credibility using the content credibility corpus. *Inf. Process. Manag.* 53, 1043–1061. doi: 10.1016/j.ipm.2017.04.003
- Kane, M. (2012). Validating score interpretations and uses. *Lang. Test.* 29, 3–17. doi: 10.1177/0265532211417210
- Karlova, N. A., and Fisher, K. E. (2013). A social diffusion model of misinformation and disinformation for understanding human information behaviour. *Inf. Res.* 18:573.
- Kingsley, K., Galbraith, G. M., Herring, M., Stowers, E., Stewart, T., and Kingsley, K. V. (2011). Why not just google it? An assessment of information literacy skills in a biomedical science curriculum. *BMC Med. Educ.* 11:1. doi: 10.1186/1472-6920-11-17
- KMK (2016). *Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in Germany. Bildung in der digitalen Welt Strategie der Kultusministerkonferenz. [Education in the digital world. KMK strategy paper]*. Retrieved from https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2018/Digitalstrategie\_2017\_mit\_Weiterbildung.pdf
- Kohnen, A. M., and Mertens, G. E. (2019). I'm always kind of double-checking: exploring the information-seeking identities of expert generalists. *Reading Res. Q.* 54, 279–297. doi: 10.1002/rrq.245
- Koltay, T. (2011). The media and the literacies: media literacy, information literacy, digital literacy. *Media Cult. Soc.* 33, 211–221. doi: 10.1177/0163443710393382
- Krämer, N. C., Preko, N., Flanagin, A., Winter, S., and Metzger, M. (2018). "What do people attend to when searching for information on the web," in *ICPS, Proceedings of the Technology, Mind, and Society Conference, Washington, DC* (New York, NY: The Association for Computing Machinery). doi: 10.1145/3183654.3183682
- Kuhlthau, C. C. (1993). A principle of uncertainty for information seeking. *J. Doc.* 49, 339–355. doi: 10.1108/eb026918
- Kuhlthau, C. C., Heinström, J., and Todd, R. J. (2008). The 'information search process' revisited: is the model still useful. *Inf. Res.* 13, 13–14.
- Lawless, K. A., Goldman, S. R., Gomez, K., Manning, F., and Braasch, J. (2012). "Assessing multiple source comprehension through evidence-centered design," in *Reaching an Understanding: Innovations in How We View Reading*

- Assessment, eds J. P. Sabatini, T. O'Reilly, and E. Albro (Lanham, MD: Rowman and Littlefield Education), 3–17.
- Leeder, C., and Shah, C. (2016). Practicing critical evaluation of online sources improves student search behavior. *J. Acad. Libr.* 42, 459–468. doi: 10.1016/j.acalib.2016.04.001
- List, A., and Alexander, P. A. (2017). Analyzing and integrating models of multiple text comprehension. *Educ. Psychol.* 52, 143–147. doi: 10.1080/00461520.2017.1328309
- Liu, O. L., Frankel, L., and Crotts Roohs, K. (2014). *Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment*. Princeton, NJ: ETS. doi: 10.1002/ets2.12009
- Lucassen, T., and Schraagen, J. M. (2011). Factual accuracy and trust in information: the role of expertise. *J. Am. Soc. Inf. Sci. Technol.* 62, 1232–1242. doi: 10.1002/asi.21545
- Lucassen, T., and Schraagen, J. M. (2013). The influence of source cues and topic familiarity on credibility evaluation. *Comput. Hum. Behav.* 29, 1387–1392. doi: 10.1016/j.chb.2013.01.036
- Maurer, A., Schloegl, C., and Dreisiebner, S. (2017). Comparing information literacy of student beginners among different branches of study. *Libellarium* 9:2. doi: 10.15291/libellarium.v9i2.280
- Maurer, M., Quiring, O., and Schemer, C. (2018). “Media effects on positive and negative learning,” in *Positive Learning in the Age of Information*, eds O. Zlatkin-Troitschanskaia, G. Wittum, and A. Dengel (Wiesbaden: Springer), 197–208. doi: 10.1007/978-3-658-19567-0\_11
- Maurer, M., Schemer, C., Zlatkin-Troitschanskaia, O., and Jitomirski, J. (2020). “Positive and negative media effects on university students' learning: preliminary findings and a research program,” in *Frontiers and Advances in Positive Learning in the Age of Information (PLATO)*, ed O. Zlatkin-Troitschanskaia (Cham: Springer International Publishing), 109–119. doi: 10.1007/978-3-030-26578-6\_8
- Mayer, R. E. (2009). *Multimedia Learning, 2nd Edn*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511811678
- Mayring, P. (2014). *Qualitative Content Analysis. Theoretical Foundation, Basic Procedures and Software Solution*. Retrieved from <https://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173>
- McCrudden, M. T., Magliano, J. P., and Schraw, G. J. (2011). *Text Relevance and Learning From Text*. Charlotte, NC: Information Age Pub. doi: 10.1007/978-1-4419-1428-6\_354
- McGrew, S., Smith, M., Breakstone, J., Ortega, T., and Wineburg, S. (2019). Improving university students' web savvy: an intervention study. *Br. J. Educ. Psychol.* 89, 485–500. doi: 10.1111/bjep.12279
- McMullin, S. L. (2018). *The correlation between information literacy and critical thinking of college students: an exploratory study* (dissertation thesis). University of North Texas, Denton, TX, United States (ProQuest LLC).
- Messick, S. (1989). “Validity,” in *Educational Measurement*, 3rd Edn., ed R. L. Linn (New York, NY: American Council on education and Macmillan), 13–104.
- Metzger, M. J. (2007). Making sense of credibility on the web: models for evaluating online information and recommendations for future research. *J. Am. Soc. Inf. Sci. Technol.* 58, 2078–2091. doi: 10.1002/asi.20672
- Metzger, M. J., and Flanagan, A. (2015). “Psychological approaches to credibility assessment online” in *The Handbook of the Psychology of Communication Technology*, ed S. S. Sundar (Chichester; Malden, MA: Wiley Blackwell), 445–466. doi: 10.1002/9781118426456.ch20
- Metzger, M. J., and Flanagan, A. J. (2013). Credibility and trust of information in online environments: the use of cognitive heuristics. *J. Pragmatics* 59, 210–220. doi: 10.1016/j.pragma.2013.07.012
- Mislevy, R. J. (2017). *Socio-Cognitive Foundations of Educational Measurement*. London: Routledge. doi: 10.4324/9781315871691
- Mislevy, R. J., and Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educ. Meas.* 25, 6–20. doi: 10.1111/j.1745-3992.2006.00075.x
- Molerov, D., Zlatkin-Troitschanskaia, O., and Schmidt, S. (2019). “Adapting the civic online reasoning assessment cross-nationally using an explicit functional equivalence approach” in *Annual Meeting of the American Educational Research Association (Toronto)*.
- Moore, T. (2013). Critical thinking: seven definitions in search of a concept. *Stud. Higher Educ.* 38, 506–522. doi: 10.1080/03075079.2011.586995
- Münchow, H., Richter, T., von der Mühlen, S., and Schmid, S. (2019). The ability to evaluate arguments in scientific texts: measurement, cognitive processes, nomological network, and relevance for academic success at the university. *Br. J. Educ. Psychol.* 89, 501–523. doi: 10.1111/bjep.12298
- Murray, M. C., and Pérez, J. (2014). Unraveling the digital literacy paradox: how higher education fails at the fourth literacy. *Issues Inf. Sci. Inf. Technol.* 11, 189–210. doi: 10.28945/1982
- Nagel, M.-T., Schäfer, S., Zlatkin-Troitschanskaia, O., Schemer, C., Maurer, M., Molerov, D., et al. (2020). How do university students' web search behavior, website characteristics, and the interaction of both influence students' critical online reasoning? *Front. Educ.* 5:1. doi: 10.3389/educ.2020.565062
- National Research Council (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: National Academies Press.
- Newman, N., Fletcher, R., Kalogeropoulos, A., and Nielsen, R. K. (2019). *Reuters Institute Digital News Report 2019. Reuters Institut for the Study of Journalism*. Available online at: [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR\\_2019\\_FINAL.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR_2019_FINAL.pdf) (accessed January 1, 2020).
- Oser, F. K. (2018). “Positive learning through negative learning - the wonderful burden of PLATO,” in *Positive Learning in the Age of Information: A Blessing or a Curse?* (Wiesbaden: Springer VS), 363–372.
- Oser, F. K., and Biedermann, H. (2020). “A three-level model for critical thinking: critical alertness, critical reflection, and critical analysis,” in *Frontiers and Advances in Positive Learning in the Age of Information (PLATO)*, ed O. Zlatkin-Troitschanskaia (Cham: Springer International Publishing), 89–106.
- Paul, R., and Elder, L. (2005). *A Guide for Educators to Critical Thinking Competency Standards, Principles, Performance Indicators, and Outcomes with a Critical Thinking Master Rubric*. Available online at: [www.criticalthinking.org](http://www.criticalthinking.org) (accessed June 24, 2020).
- Paul, R., and Elder, L. (2008). *The Thinker's Guide for Conscientious Citizens on How to Detect Media Bias and Propaganda in National and World News: In National and World News, 4th Edn*. Dillon Beach, CA: The Foundation for Critical Thinking.
- Pellegrino, J. W. (2017). “Teaching, learning and assessing 21st century skills,” in *Educational Research and Innovation. Pedagogical Knowledge and the Changing Nature of the Teaching Profession*, ed S. Guerriero (Paris: OECD Publishing), 223–251. doi: 10.1787/9789264270695-12-en
- Pernice, K. (2017). *F-Shaped Pattern of Reading on the Web: Misunderstood, but Still Relevant (Even on Mobile)*. *World Leaders in Research-Based User Experience*. Available online at: <https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/> (accessed June 24, 2020).
- Pirolli, P., and Card, S. (1999). Information foraging. *Psychol. Rev.* 106, 643–675. doi: 10.1037/0033-295X.106.4.643
- Podgornik, B. B., Dolničar, D., and Glažar, S. A. (2017). Does the information literacy of university students depend on their scientific literacy? *Eurasia J. Math. Sci. Technol. Educ.* 13, 3869–3891. doi: 10.12973/eurasia.2017.00762a
- Powers, E. M. (2019). How students access, filter and evaluate digital news: choices that shape what they consume and the implications for news literacy education. *J. Lit. Technol.* 20:3.
- Reese, S. D., and Shoemaker, P. J. (2016). A media sociology for the networked public sphere: the hierarchy of influences model. *Mass Commun. Soc.* 19, 389–410. doi: 10.1080/15205436.2016.1174268
- Rieh, S. Y. (2010). “Credibility and cognitive authority of information” in *Encyclopedia of Library and Information Sciences*. 1, 1337–1344.
- Rieh, S. Y. (2014). Credibility assessment of online information in context. *J. Inf. Sci. Theory Pract.* 2, 6–17. doi: 10.1633/JISTaP.2014.2.3.1
- Roozenbeek, J., and van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Commun.* 5:133. doi: 10.1057/s41599-019-0279-9
- Rouet, J. F. (2006). *The Skills of Document Use: From Text Comprehension to Web-Based Learning*. Mahwah, NJ: Erlbaum. Available online at: <http://www.loc.gov/catdir/enhancements/fy0625/2005052083-d.html> (accessed June 24, 2020). doi: 10.4324/9780203820094
- Salmerón, L., Cañas, J. J., Kintsch, W., and Fajardo, I. (2005). Reading strategies and hypertext comprehension. *Discourse Process.* 40, 171–191. doi: 10.1207/s15326950dp4003\_1

- Salmerón, L., Kammerer, Y., and García-Carrión, P. (2013). Searching the web for conflicting topics: page and user factors. *Comput. Hum. Behav.* 29, 2161–2171. doi: 10.1016/j.chb.2013.04.034
- Samson, S. (2010). Information literacy learning outcomes and student success. *J. Acad. Libr.* 36, 202–210. doi: 10.1016/j.acalib.2010.03.002
- Sanders, L., Kurbanoglu, S., Boustany, J., Dogan, G., and Becker, P. (2015). Information behaviors and information literacy skills of LIS students: an international perspective. *J. Educ. Libr. Inf. Sci. Online* 56, 80–99. doi: 10.12783/issn.2328-2967/56/S1/9
- Schmidt, S., Zlatkin-Troitschanskaia, O., Roeper, J., Klose, V., Weber, M., Bültmann, A.-K., et al. (2020). Undergraduate students' critical online reasoning - process mining analysis. *Front. Psychol.* (in press). doi: 10.3389/fpsyg.2020.576273
- Schnell, R., Hill, P. B., and Esser, E. (2011). *Methoden der empirischen Sozialforschung [Methods of Empirical Social Research]*, 9th Edn. Munich: Oldenbourg.
- Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., and Menczer, F. (2017). *The Spread of Fake News by Social Bots*. Available online at: <https://arxiv.org/abs/1707.07592> (accessed June 24, 2020).
- Shavelson, R. J., Zlatkin-Troitschanskaia, O., Beck, K., Schmidt, S., and Mariño, J. P. (2019). Assessment of university students' critical thinking: next generation performance assessment. *Int. J. Test.* 19, 337–362. doi: 10.1080/15305058.2018.1543309
- Shavelson, R. J., Zlatkin-Troitschanskaia, O., and Mariño, J. (2018). "International performance assessment of learning in higher education (iPAL): research and development," in *Assessment of Learning Outcomes in Higher Education – Cross-National Comparisons and Perspectives*, eds O. Zlatkin-Troitschanskaia, M. Toepper, H. A. Pant, and C. Lautenbach (Wiesbaden: Springer), 193–214. doi: 10.1007/978-3-319-74338-7\_10
- Shoemaker, P. J., and Reese, S. D. (2014). *Mediating the Message in the 21st Century: A Media Sociology Perspective*, 3rd Edn. New York, NY: Routledge/Taylor and Francis Group. doi: 10.4324/9780203930434
- Snow, C. E. (2002). *Reading for Understanding: Toward an RandD Program in Reading Comprehension*. Santa Monica CA: Rand.
- Solano-Flores, G., Backhoff, E., and Contreras-Niño, L. Á. (2009). Theory of test translation error. *Int. J. Test.* 9, 78–91. doi: 10.1080/15305050902880835
- Sparks, J. R., Katz, I. R., and Beile, P. M. (2016). Assessing digital information literacy in higher education: a review of existing frameworks and assessments with recommendations for next-generation assessment. *ETS Res. Rep. Ser.* 2016, 1–33. doi: 10.1002/ets2.12118
- Stanovich, K. E., West, R., and Toplak, M. E. (2016). *The Rationality Quotient: Toward a Test of Rational Thinking*. Cambridge, MA: The MIT Press. doi: 10.7551/mitpress/9780262034845.001.0001
- Sundar, S. S. (2008). "The MAIN model: a heuristic approach to understanding technology effects on credibility," in *Digital Media, Youth, and Credibility*, eds M. J. Metzger, and A. J. Flanagin (Cambridge: MIT Press), 73–100.
- Tanaka, K. (2009). "Web search and information credibility analysis: bridging the gap between web1.0 and web2.0," in *ICUIMC 2009: Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication* (Suwon), 39–44.
- Tanaka, K., Kawai, Y., Zhang, J., Nakajima, S., Inagaki, Y., Ohshima, H., et al. (2010). "Evaluating credibility of web information," in *Proceedings of the 4th International Conference on Ubiquitous Information Management and Communication - ICUIMC '10*, eds W. Kim, D. Won, K.-H. You, and S.-W. Lee (New York, NY: ACM Press), 1–10. doi: 10.1145/2108616.2108645
- Taylor, A., and Dalal, H. A. (2014). Information literacy standards and the world wide web: results from a student survey on evaluation of Internet information sources. *Inf. Res.* 19:4.
- Threadgill, E. J., and Price, L. R. (2019). Assessing online viewing practices among college students. *J. Media Lit. Educ.* 11, 37–55. doi: 10.23860/JMLE-2019-11-2-3
- Toplak, M. E., Liu, E., MacPherson, R., Toneatto, T., and Stanovich, K. E. (2007). The reasoning skills and thinking dispositions of problem gamblers: a dual process taxonomy. *J. Behav. Decis. Mak.* 20, 103–124. doi: 10.1002/bdm.544
- Toulmin, S. (2003). *The Uses of Argument, Updated Edn*. Cambridge, NY: Cambridge University Press. doi: 10.1017/CBO9780511840005
- Tseng, S., and Fogg, B. J. (1999). Credibility and computing technology. *Commun. ACM* 42, 39–44. doi: 10.1145/301353.301402
- Van Eemeren, F. H. (2013). Fallacies as derailments of argumentative discourse: acceptance based on understanding and critical assessment. *J. Pragmatics* 59, 141–152. doi: 10.1016/j.pragma.2013.06.006
- Walraven, A., Brand-Gruwel, S., and Boshuizen, H. P. A. (2008). Information-problem solving: a review of problems students encounter and instructional solutions. *Comput. Hum. Behav.* 24, 623–648. doi: 10.1016/j.chb.2007.01.030
- Walraven, A., Brand-Gruwel, S., and Boshuizen, H. P. A. (2009). How students evaluate information and sources when searching the world wide web for information. *Comput. Educ.* 52, 234–246. doi: 10.1016/j.compedu.2008.08.003
- Walton, D. (2006). *Fundamentals of Critical Argumentation. Critical Reasoning and Argumentation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807039
- Walton, D. (2017). Value-based argumentation in mass audience persuasion dialogues. *COGENCY* 9, 139–159.
- Walton, D., Reed, C., and Macagno, F. (2008). *Argumentation Schemes*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511802034
- Walton, G., Barker, J., Pointon, M., Turner, M., and Wilkinson, A. (2020). "Information literacy and the societal imperative of information discernment," in *Informed Societies: Why Information Literacy Matters for Citizenship, Participation and Democracy*, ed S. Goldstein (London: Facet Publishing), 149. doi: 10.29085/9781783303922.010
- Wathen, C. N., and Burkell, J. (2002). Believe it or not: factors influencing credibility on the web. *J. Am. Soc. Inf. Sci. Technol.* 53, 134–144. doi: 10.1002/asi.10016
- Weekley, J. A., and Ployhart, R. E. (2013). *Situational Judgment Tests: Theory, Measurement, and Application*. Mahwah: Erlbaum.
- Wierzbicki, A. (2018). *Web Content Credibility*. New York, NY: Springer Berlin Heidelberg. doi: 10.1007/978-3-319-77794-8
- Wineburg, S., Breakstone, J., McGrew, S., and Ortega, T. (2018). "Why google can't save us. The challenges of our post-gutenberg moment," in *Positive Learning in the Age of Information*, eds O. Zlatkin-Troitschanskaia, G. Wittum, and A. Dengel (Wiesbaden: Springer), 221–228. doi: 10.1007/978-3-658-19567-0\_13
- Wineburg, S., and McGrew, S. (2016). Why students can't google their way to the truth: fact-checkers and students approach websites differently. *Educ. Week* 36, 22–28.
- Wineburg, S., and McGrew, S. (2017). *Lateral Reading: Reading Less and Learning More When Evaluating Digital Information (Working Paper)*. Available online at: <https://ssrn.com/abstract=3048994> (accessed July 22, 2018). doi: 10.2139/ssrn.3048994
- Wineburg, S., McGrew, S., Breakstone, J., and Ortega, T. (2016a). Evaluating information: the cornerstone of civic online reasoning. *Stanford Digital Repository*.
- Wineburg, S., McGrew, S., Breakstone, J., and Ortega, T. (2016b). *Evaluating Information: The Cornerstone of Civic Online Reasoning: Executive summary*. Stanford History Education Group. Available online at: <http://purl.stanford.edu/fv751yt5934> (accessed June 24, 2020).
- Winter, S., Metzger, M. J., and Flanagin, A. J. (2016). Selective use of news cues: a multiple-motive perspective on information selection in social media environments. *J. Commun.* 66, 669–693. doi: 10.1111/jcom.12241
- Xie, I. (2008). *Interactive Information Retrieval in Digital Environments*. Hershey: IGI Global. doi: 10.4018/978-1-59904-240-4
- Zhang, S., and Duke, N. K. (2008). Strategies for internet reading with different reading purposes: a descriptive study of twelve good internet readers. *J. Lit. Res.* 40, 128–162. doi: 10.1080/10862960802070491
- Zhang, S., Duke, N. K., and Jiménez, L. M. (2011). The WWWDOT approach to improving students' critical evaluation of websites. *Reading Teach.* 65, 150–158. doi: 10.1002/TRTR.01016
- Zlatkin-Troitschanskaia, O. (2020). *Frontiers and Advances in Positive Learning in the Age of InformaTiOn (PLATO)*. Cham: Springer International Publishing. doi: 10.1007/978-3-030-26578-6
- Zlatkin-Troitschanskaia, O., Beck, K., Fischer, J., Braunheim, D., Schmidt, S., and Shavelson, R. J. (2020a). The role of students' beliefs when critically reasoning from multiple contradictory sources of information in performance assessments. *Front. Psychol.* 11:2192. doi: 10.3389/fpsyg.2020.02192
- Zlatkin-Troitschanskaia, O., Brückner, S., Molerov, D., and Bisang, W. (2020b). "What can we learn from theoretical considerations and empirical evidence on learning in higher education? Implications for an interdisciplinary research framework," in *Frontiers and Advances in Positive Learning in the Age*

- of InformaTiOn (PLATO)*, ed. O. Zlatkin-Troitschanskaia (Cham: Springer International Publishing), 287–309.
- Zlatkin-Troitschanskaia, O., Dengel, A., and Wittum, G. (2018a). *Positive Learning in the Age of Information: A Blessing or a Curse?* Wiesbaden: Springer VS. doi: 10.1007/978-3-658-19567-0
- Zlatkin-Troitschanskaia, O., Schmidt, S., Molerov, D., Shavelson, R. J., and Berliner, D. (2018). "Conceptual fundamentals for a theoretical and empirical framework of positive learning," in *Positive Learning in the Age of Information: A Blessing or a Curse?*, eds O. Zlatkin-Troitschanskaia, A. Dengel, and G. Wittum (Wiesbaden: Springer VS.), 29–50.
- Zlatkin-Troitschanskaia, O., Shavelson, R. J., Schmidt, S., and Beck, K. (2019). On the complementarity of holistic and analytic approaches to performance assessment scoring. *Br. J. Educ. Psychol.* 89, 468–484. doi: 10.1111/bjep.12286
- Zlatkin-Troitschanskaia, O., Toepper, M., Molerov, D., Buske, R., Brückner, S., Pant, H. A., et al. (2018b). "Adapting and validating the collegiate learning assessment to measure generic academic skills of students in Germany: implications for international assessment studies in higher education," in *Assessment of Learning Outcomes in Higher Education*, eds O. Zlatkin-Troitschanskaia, M. Toepper, H. A. Pant, C. Lautenbach, and C. Kuhn (Cham: Springer International Publishing), 245–266. doi: 10.1007/978-3-319-74338-7\_12
- Zumbo, B. D., and Hubley, A. M. (2017). Understanding and Investigating Response. *Processes in Validation Research*. Cham: Springer, 69. doi: 10.1007/978-3-319-56129-5
- Zylka, J., Christoph, G., Kröhne, U., Hartig, J., and Goldhammer, F. (2015). Moving beyond cognitive elements of ICT literacy. First evidence on the structure of ICT engagement. *Comput. Hum. Behav.* 53, 149–160. doi: 10.1016/j.chb.2015.07.008

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Molerov, Zlatkin-Troitschanskaia, Nagel, Brückner, Schmidt and Shavelson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.