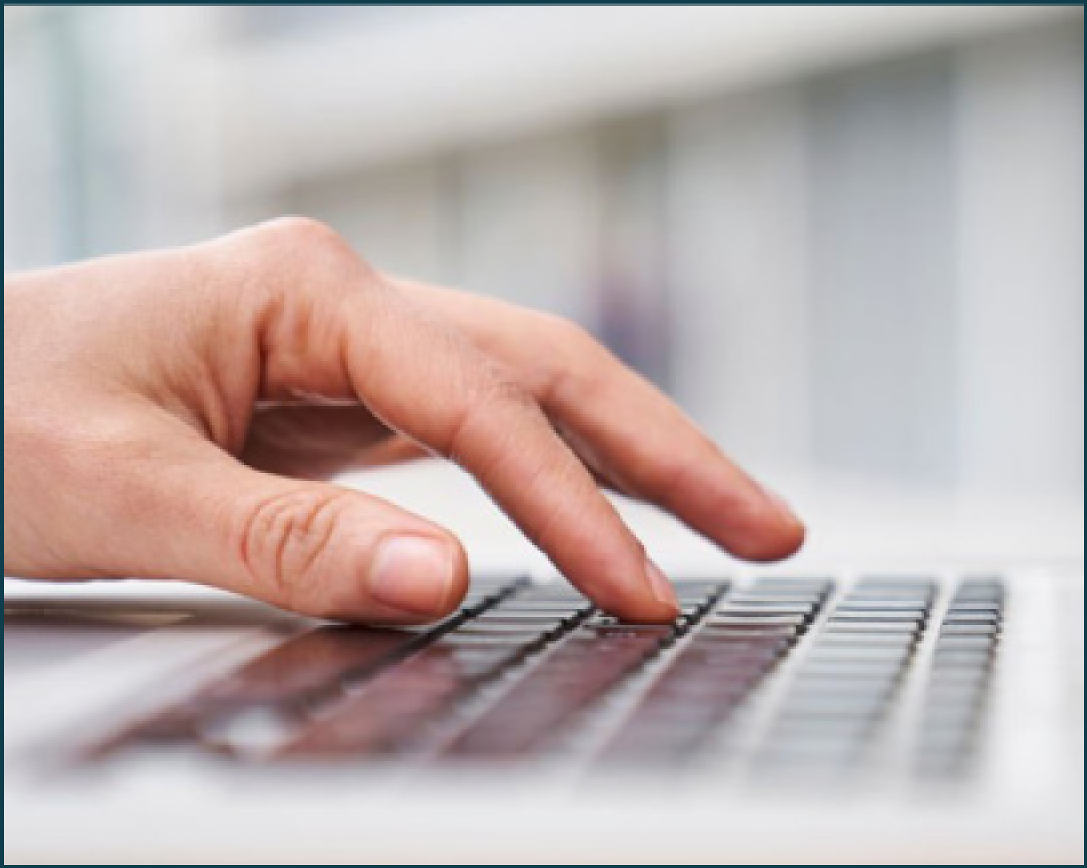
*ets

August
2024

Opportunities and Challenges for Assessing Digital and AI Literacies



AUTHORS

Jesse R. Sparks,
Managing Senior
Research Scientist

Teresa M. Ober,
Research Scientist

Caitlin Tenison,
Research Scientist

Burcu Arslan,
Research Scientist

Ido Roll,
Senior
Research Director

Paul Deane,
Principal
Research Scientist

Diego Zapata Rivera,
Distinguished
Presidential Appointee

Reginald M. Gooch,
Research
Project Manager

Tenaha O'Reilly,
Managing Principal
Research Scientist

Contents

Contents

Executive summary

The digital landscape in which we access, process, share, create and communicate information continuously evolves, as do the skills we need to interact within these digital environments to achieve our goals.

The rise of generative AI (GenAI) introduces even greater changes to how we research, synthesize and communicate information in digital contexts. GenAI is a type of AI that uses generative models trained on vast amounts of data that specialize in generating different types of content, including images, code or text. When powered by large language models (LLMs) specifically trained on text data, GenAI models can generate content based on natural language input.

A new understanding of the essential skills required to be literate in digital environments that involve using and interacting with AI systems is needed. This is especially challenging, as the changing nature of AI creates a moving target.

- In a time of rapid technological change, utilizing technology requires evolving understandings and practices.
- Thus, we must understand how to assess these evolving competencies. Many emerging "AI literacy" frameworks attempt to define what learners should know and do while engaging with AI technologies.
- Such frameworks generally extend traditional concepts of digital literacies to contexts involving AI by articulating the types of knowledge, skills and other attributes learners need to define, access, evaluate, manage, integrate and create informational content using AI systems in these contexts.

To understand digital and AI literacies, we must assess not only the end product(s) that reflect one's ability to use AI to gather, organize, critically evaluate and express information, but also the processes used to achieve those ends, including:

- Knowing how to use AI technologies effectively to solve problems (such as engaging with and refining prompts over time within threads of conversation) and communicate ideas; Recognizing the use of AI in digital environments and understanding that AI responses may be inaccurate; Knowing how to evaluate and verify information produced with AI tools.
- AI ethics should also be assessed: the ability to understand and reason about factors impacting the access and use of information such as biases, privacy, and data ownership.

The digital divide (unequal access to technology) must be considered when assessing digital and AI literacies, along with variations in the cultural norms and values reflected in the assessment, particularly in large-scale international contexts.

Assessment and teaching of digital and AI literacies are a matter of equity. What people know and can do affects how vulnerable they are to unethical uses of technologies.

Opportunities and Challenges for Assessing Digital and AI Literacies

In nearly all facets of daily life, we increasingly face challenges posed by a continuously evolving digital media landscape that introduces and affords new ways of accessing, processing, sharing, creating and communicating information to achieve our goals.

Ongoing and emerging advances in artificial intelligence (AI) technologies have further complicated and disrupted our interactions with digital media, transforming and enhancing those interactions in ways that continually evolve as technological development proceeds.[1]

These ever-evolving contexts, tools and applications introduce opportunities and challenges for how we conceptualize — and how we assess — important constructs such as digital literacies as they are reflected in contemporary use of these emerging technologies.[2] Widespread public use of AI has great potential to introduce changes to how we enact digital literacies, including how we conduct inquiry, how we synthesize information and how we communicate our ideas.

Digital assessments must also evolve to reflect these realities, to understand and support claims about individuals' proficiency with such evolving competencies, as well as their preparedness for participation in 21st century life, including informed citizenship and productive employment in this AI-enabled future.

In this report, we address the question of how to conceptualize and assess these crucial digital and AI literacies — especially in the context of recent advances in generative AI — and the potential risks and consequences of such assessment.

Our aim is to provide readers with guidance to inform the design of scenarios and tasks that can be used to assess digital literacies in contexts that meaningfully incorporate the use of AI tools.

We hope this guidance will enable both assessment developers and users of assessment scores to make claims and draw inferences about learners' proficiency with ethical and appropriate use of AI-enabled technologies in the service of completing information-based digital literacies tasks.

The structure of this report is as follows:

- Section 1 discusses how to conceptualize these important AI literacies within the larger context of digital literacies and as the next evolutionary step in the development of literacies more generally.
- Section 2 discusses new affordances and challenges of AI for assessing digital literacies and provides an illustrative example in the context of a web-based inquiry scenario that incorporates such technologies.
- Section 3 provides recommendations for assessment development, including principled design approaches and contextual considerations, using the scenario in Section 2 to provide concrete examples where appropriate.
- Section 4 discusses some of the risks and consequences of creating and administering such digital and AI literacies assessments at a global scale.
- Section 5 offers some key conclusions.

SECTION 01.

Conceptualizing digital and AI literacies

Digital literacies have historically been broadly defined, reflecting the wide range of knowledge, skills and other attributes (KSAs) that can be deployed while completing information-based tasks in digital environments. Gilster first coined the term “digital literacy” in 1997 and focused on the multimodal nature of digital media and how the emergence of the Internet required new skills for searching and retrieving information.[3]

By using the term “digital literacies,” we emphasize the multidimensional, complex and diverse nature of digital literacies practices[4] and the “tightly intertwined,” “interdependent” and “integrated” nature of the meaning-making processes involved in accessing, reading, writing, navigating, creating and communicating digital content when working with digital texts, tools, technologies, interfaces and networks.[5] This perspective emphasizes both the new forms of literacies afforded by emerging digital technologies and the inherently social, contextual and cultural aspects of these digital literacies practices within digital spaces.[6]

Recent accounts of digital literacies from a 21st century skills perspective are consistent with these perspectives in emphasizing that successfully accessing, navigating and using digital information requires both foundational communication, critical thinking, problem-solving, collaboration, creativity, information management and technical competencies, as well as contextual skills focused on ethical and cultural awareness, flexibility, self-direction and lifelong learning.[7]

THE EVOLUTION OF SKILLS NEEDED TO ACCESS, SYNTHESIZE AND SHARE INFORMATION

Pre-1990s	1990s	2020s
Traditional Literacies	**Digital Literacies**	**AI Literacies**
Skills needed to understand and communicate, including reading and writing	Skills required to access, manage and evaluate digital information	Skills required for accessing, managing and creating information via AI
Prior to the advent of the Internet, information was disseminated via printed texts, books, reports and journals.	Information began to be stored, shared and accessed via the Internet and emerging web search tools.	Enabled by AI, accessing and generating information is much faster but may be less reliable.

Digital literacies have been widely assessed. In previous work, ETS researchers developed a comprehensive operational definition for assessment purposes based on syntheses of frameworks and existing assessments used in K-12 and higher education contexts.[8]

Digital information literacy was defined as "the ability to function in a knowledge society through the appropriate use of information and communication technology to solve information problems, including the ability to research, organize and synthesize information through digital technology and having a fundamental understanding of the ethical/legal issues surrounding the use of such information."[9]

This framework identified six key competencies to be assessed, including defining, accessing, evaluating, managing, integrating and creating information, highlighting the multifaceted nature of digital literacies and the important role of inquiry, critical thinking and information-based problem-solving within this larger construct.[10]

This framework also emphasized the process-driven nature of digital literacies tasks and recommended use of digital performance-based assessments to measure these competencies. Such digital assessments enable assessment developers to go beyond knowledge and skills to measure the processes of using digital technologies, in addition to the products of such use, and to incorporate evidence of those processes into scoring and reporting.

Emerging work by ETS researchers has revisited the construct of digital literacies with greater emphasis on ethical and socially responsible use of technologies. This work defined digital literacies as "a set of knowledge, skills and attitudes necessary to use digital technologies and tools productively and responsibly across social, academic and professional settings."[11]

This perspective frames digital literacies as a highly contextualized set of competencies that are situated and enacted within specific settings, tasks and sociocultural contexts.

Relevant competencies include the ability to access, manage, and understand digital information, to communicate and collaborate via digital technologies, to create digital content and to engage constructively and responsibly in digital environments. The contexts in which these competencies are applied shape both how this complex construct is conceptualized (e.g., in educational and assessment frameworks) and how individuals make meaning, engage in critical evaluation, and create information products in continually shifting, dynamic digital media environments that often require integration of different modalities (e.g., audio, video, text), asynchronous and synchronous communication and interactivity (i.e., offline vs. real-time), multitasking (i.e., task switching), critical evaluation and synthesis of information from multiple sources of varied quality and other emerging complexities.[12]

This perspective acknowledges that digital literacies are inseparable from the social, cultural and technological contexts of their use, and therefore, conceptions of digital literacies must also continually evolve to keep pace with these rapidly evolving contexts.[13]

While competencies reflecting digital literacies remain important targets for assessment, emerging and rapid advancements in AI — especially GenAI — may have implications for how those competencies are understood and applied in practice when interacting with AI-enabled tools. Next, we describe how to conceptualize these advancements and their implications for digital literacies.

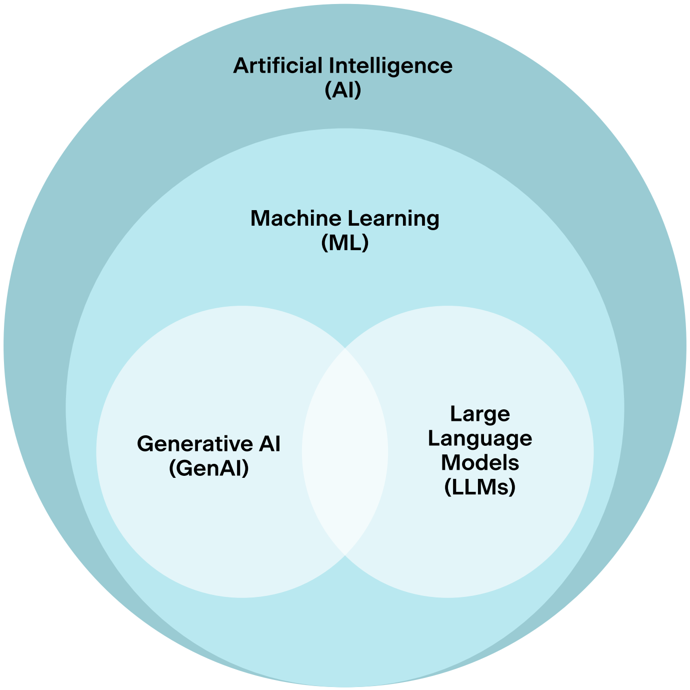
Implications of AI for digital literacies

The Organisation for Economic Co-operation and Development (OECD) defines an *AI system* as "*a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments.*"[14] This definition is broad and encompasses several different types of AI systems, including machine learning (ML). ML is a branch of AI that uses computational algorithms to learn from experience and improve its analyses.[15] These algorithms process large datasets to recognize patterns, enabling the machine to make autonomous decisions or recommendations and refine its predictions through iterations.

There are various types of ML algorithms. *GenAI* is a form of ML that uses generative models trained on vast amounts of data and specializes in generating content, including images, video, music and code. *Large language models* (LLMs) are specifically trained on text data and can perform both generative tasks, such as summarizing text and answering questions, and nongenerative tasks, such as classification of text into various categories. When powered by LLMs, GenAI models can create content based on natural language input. Thus, GenAI and LLMs are examples of ML approaches within the larger space of AI systems (see Figure 1).

There is growing awareness that AI systems have and will continue to have consequences for teaching, learning and assessment in formal educational settings such as K-12 schools as well as colleges and universities.

Figure 1. A visual representation of the relationships among the AI-related terms used in this report.



The U.S. Department of Education recognizes that such technologies have great potential to advance educational goals, yet stress the need for users to critically evaluate and limit uses of such technologies, particularly considering their rapid pace of development.[16] The speed of such advancements — and the technical knowledge required to engage with AI technologies critically and responsibly — also raises the challenge of how to identify and assess what learners should know and be able to do with respect to such technologies.

For example, a 2023 Pew survey of U.S. adults found that only 30% of respondents demonstrated high awareness of the influence of AI in commonly used digital tools.[17]

Respondents with higher educational attainment and income tended to report greater awareness of the influence of AI, illustrating possible disparities in access, use or understanding of AI and its consequences.

In contrast, many children today may already have some awareness of AI's influence on their daily lives. One study found that 5th and 6th grade students already hold naïve concepts about AI.[18] Despite such preconceptions, at present, efforts to systematically identify and assess the most important competencies related to digital literacies and AI technologies learners need to thrive in education and careers are limited.

Efforts to identify critical competencies related to the use of digital media and AI technologies are not only necessary to inform assessment development but are also essential in preparing learners for future success through relevant, practical instruction and career preparation. There is growing recognition that AI technologies have disrupted, and will continue to disrupt, the workforce.[19] Certain occupations may be particularly susceptible to automation driven by AI.[20]

Individuals who can develop the necessary foundational digital competencies and effectively use these competencies to navigate the future of work — that is, using AI synergistically and in complement to human reasoning and decision-making — are likely to increase the durability of their skills in a potentially volatile workforce.[21] This view also emphasizes the need to define concrete and attainable competencies that can help people utilize AI to achieve their goals.

Given the ubiquity of AI technologies in daily activities and the projected impact on future career opportunities, it is imperative for education to define the critical competencies and related assessment strategies.

Such efforts could help to identify gaps in learners' knowledge and skills and ultimately provide support for them to engage critically and responsibly with digital technologies that use AI.[22] Many emerging frameworks attempt to define what learners should know and do while engaging with AI, often under the heading of "AI literacy."[23]

Such frameworks generally extend traditional concepts of digital literacies to contexts involving AI by articulating the KSAs learners need to define, access, evaluate, manage, integrate and create informational content using AI tools in the context of these digital literacies processes and practices.[24]

These frameworks differ somewhat in emphasis depending on their purpose, target population and domain of study, with some emphasizing technical knowledge and skills needed to build AI models[25] and others emphasizing the sociotechnical and ethical implications of AI within educational systems.[26]

Common themes among such frameworks that may serve as useful targets for assessment include accessing, managing and evaluating information; creating, sharing and communicating with digital content and secure and ethical use of information. A recent systematic literature review identified six broad constructs of focus for AI literacy: know and understand, use and apply, evaluate, navigate ethically, recognize and create.[27]

Given our goal to situate AI literacy within the larger context of digital literacies, we focus the remainder of our discussion on four broad constructs (defined in Table 1) most relevant to digital literacies. These constructs capture AI literacy skills important to nontechnical users.

TABLE 1. AI literacy constructs[28]

AI Literacy Construct	Description
Recognize and understand	The ability to identify the presence and role of AI in various digital tools and contexts. This includes an awareness or understanding of how AI can influence how information is presented and transmitted.
Use and apply	This construct involves the practical skills required to interact effectively with AI technologies and apply them to solve problems, create content, communicate and collaborate.
Evaluate	Evaluation skills involve critically assessing AI technology. This includes both evaluating the reliability and accuracy of the outputs of AI systems and the suitability for these systems use for specific tasks.
Navigate ethically	This construct focuses broadly on understanding the ethical implications of AI technologies and their applications and demonstrating ethical access and use of digital information.

AI literacies: Evolution, not revolution

Conceptualizations of AI literacies in digital environments can be considered within broader historical developments in how we conceptualize literacies. How do these emerging AI literacies affect our understanding of what it means to be "literate" in the 21st century? In some regards, the emergence of AI is the next chapter in a long history of how reading and writing have developed with changes in technology.

In the course of human history, the use of oral language preceded the widespread use of print literacy. Print literacy involves the processes of decoding and comprehension[29] as well as the cognitive resources to seek and acquire new information from print sources.

In more recent decades within the so-called "information age,"[30] the construct of digital literacy emerged to encompass various literacies within the context of digitally mediated environments, such as online reading/writing, web searching, use of news media, interacting on social media and so on.

However, advances in AI, including GenAI and LLMs, are redefining not only what it means to be literate, but are also changing the nature of the texts and information sources to be created, used and evaluated.

As information technologies continue to evolve in light of advances in AI, so too do the key literacies required for engaging with and using information.

For example, there appears to be a potential shift from understanding, using and creating long, cohesive, static and reliable single sources, to short, often contradictory, dynamic, interactive and multiple questionable sources. These trends can be seen in the development of literacy technologies from books, through the World Wide Web and to today's GenAI.

These changes bring about considerable disruptions. For example, some have argued that knowledge retention (i.e., memory and recall) likely mattered considerably more when synchronous spoken exchanges were a primary means of receiving and expressing ideas.[31] Innovations in information technology reduced many barriers related to seeking and accessing new information and thus enabled exposure to vast quantities of information, in multiple modalities, from a wider range of sources.[32] Overall, however, the required literacy competencies seem to have become more sophisticated — from recall to sense-making to synthesis and critical evaluation. This increase in complexity entails increasingly complex literacy assessments.[33]

Recent advances in AI appear to be driving further evolution. Though previous information technologies had largely solved the problem of information transfer, creating coherent content (such as in the form of written text) was still a time-consuming process. In particular, the advent of GenAI has now made content creation (including creation of text, images and videos) a relatively easy process, assuming an individual has the basic skills necessary to craft an appropriate prompt.

However, given the ease and widespread use of this content creation, there remains a need to prepare individuals to identify accurate and reliable content and to use the tools to generate content in ways that do not compromise the individual's potential to make responsible and autonomous decisions.

Given the relative novelty of GenAI as a tool for public use, there is pressing demand to identify and evaluate its affordances and what potentially new or emerging KSAs are required for individuals to thrive in these contexts.

To identify the literacy skills necessary to thrive in the present era, it may be helpful to consider the role of traditional reading and writing skills for academic and career success. Traditional literacy skills have been critical in enabling individuals to process information quickly and easily. Efficiency of comprehension ensures that an individual can process larger amounts of more complex content. The ability to comprehend text also depends on certain foundational skills such as decoding, reading fluency, vocabulary, linguistic knowledge and having sufficient prior content knowledge.[34]

Similar issues arise with writing; for example, fluency enables the higher level thinking that is required when individuals seek to integrate and express information from a variety of sources through synthesis. Foundational reading and writing skills are further critical to enable individuals to apply self-regulated processes such as reviewing, reflecting, planning, revising, rethinking and reworking, that enable them to understand and elaborate on ideas gleaned through text or other information sources.[35]

Yet while the need for strong literacy skills is widely recognized, many adult learners still struggle with basic aspects such as decoding, fluency and simple reading comprehension (e.g., locate information, extract main ideas, draw minor inferences).[36] Such skills are even more important when the accuracy of information also must be critically evaluated, synthesized and expressed for specific purposes because it is difficult to apply these higher level skills if the basic message is not decoded, and subsequently, comprehended.

Literacies related to information-seeking, comprehension and written communication are likely already being affected by the widespread use of LLMs.

The use of LLMs appears to have created new affordances, each with a unique set of associated opportunities and challenges.

First, LLMs have led to further proliferation of information, and consequently, a greater imperative for knowing how to evaluate and synthesize such vast amounts of information.[37] Consider that the outputs of AI often seem very plausible for less knowledgeable readers, when in fact, they may be false or totally fabricated, a phenomenon referred to as a “hallucination.”[38] These plausible yet erroneous outputs could potentially exacerbate the spread of mis- and disinformation and cause harm.[39]

Second, LLMs have considerably reduced barriers and levels of requisite skills needed to synthesize and create seemingly coherent content.[40]

Third, LLMs have led to an increase in the possibilities for the frequency, quality and types of interactivity,[41] which may potentially benefit learners in terms of promoting engagement, but cause harm in instances where false or misleading information is spread or reinforced via such exchanges.

Given that GenAI provides such affordances, there is tension between what skills an individual needs to acquire (e.g., to develop deep understanding, to have agency and control, etc.) and what digital literacies tasks they may want to accomplish with relative ease with the aid of AI tools.

Assessing specific digital and AI literacies competencies is crucial to document, understand and support the development of relevant KSAs. Assessment of digital and AI literacies can provide educators and policymakers with information about whether learners are prepared to use AI in productive, meaningful ways. Such assessment must be done with an eye toward the larger question that should be pursued — how can individuals be enabled to gather, organize, critically evaluate and express information for themselves with support of AI-enabled technologies?

SECTION 02.

Affordances and challenges of assessing digital and AI literacies

Assessments of digital literacies have frequently utilized simulated web environments to create realistic contexts where individuals can demonstrate their digital literacy skills.[42] These assessments typically involve complex information-based tasks such as searching for specific information or preparing a presentation, requiring participants to navigate and utilize simulated digital tools effectively in the context of conducting an inquiry.

From a digital literacies perspective, such digital inquiry tasks can be considered in terms of a multi-step, iterative process beginning with understanding the search task, locating and selectively accessing content, assessing the relevance and quality of results, processing the contents of those individual results and encoding or extracting key details, through to a synthesis of the information into a complete answer to the inquiry question that can then be communicated to various audiences.[43]

Next, we consider how digital inquiry tasks using simulated web search tools could be updated to also further provide opportunities for assessing important AI literacy constructs (see Table 1) by incorporating LLM-enabled web search. While the nuances of assessment design will be discussed in Section 3, the purpose of the illustrative example in this section is to demonstrate how digital and AI literacies constructs might overlap and mutually support completion of this kind of digital inquiry task (see Figure 2).

AI impacts each part of the inquiry cycle, introducing dynamic challenges for the measurement of digital literacy skills.

Figure 2. Digital and AI literacy constructs involved in completing digital inquiry tasks.

To effectively assess AI literacy, it's essential to evaluate not just the end products, but the processes that reflect one's ability to use AI to gather, organize, critically evaluate and express information.

Digital inquiry tasks can be used to measure both digital and AI literacies in an integrated fashion.

DIGITAL INQUIRY

Digital literacies supporting digital inquiry involve a cyclical process of defining problems and information needs; locating sources, evaluating sources, processing, analyzing and synthesizing those sources and communicating results.

AI LITERACIES CONSTRUCTS

Use and Apply AI

Interact effectively with AI technologies and apply them to solve problems and communicate ideas by engaging with and refining prompts within conversational threads.

Recognize and Evaluate AI

Identify the presence and role of AI in various digital tools and contexts, understanding how AI can influence the ways information is presented and the potential inaccuracies in AI responses.

Know how to evaluate and verify information produced with AI tools.

Navigate AI Ethically

Understand the ethical implications of AI technologies and their applications, demonstrating the ability to reason about biases, privacy, data ownership and societal impact.

Figure 2 illustrates how key AI literacies constructs (as shown in concentric semicircles) overlap with distinct digital literacies constructs involved in digital inquiry tasks (as shown in the outer circle reflecting an inquiry cycle).

Questions that should be considered in designing a digital and AI literacies assessment that measures learners' ability to complete this type of digital inquiry task may include:

- What are productive questions that learners can ask? That is, do they have the vocabulary, knowledge and control of strategies to figure out how to interact effectively with the AI to get the information they need in this context?
- How well do learners understand that the GenAI may not itself give accurate summaries when used for synthesis purposes? Do they know how to verify the accuracy of such GenAI responses? Do they have the knowledge and skills to evaluate AI summaries when they may sound very convincing even when wrong (i.e., in the event of GenAI hallucinations)?
- How does the use of GenAI in a search context affect the likelihood of critically reading, evaluating sources or checking for plagiarism against accurate and reliable sources of information in sufficient depth?
- How do learners build mental models they will remember based on their searches and enable them to use what they have learned to suit their own purposes?
- How well do learners understand biases and other social factors that affect the outcomes of their search with AI?

Illustrative example: A web-based digital inquiry task

Web-based digital inquiry offers a compelling scenario to illustrate how AI literacy fits within and supports a broader framework of digital literacies. From its inception, web search has been driven by advancements in AI. These advancements can be seen directly in improvements in algorithms supporting information retrieval, which have evolved alongside efforts to represent and optimize the availability of information on websites and other digital sources.

Beyond information retrieval, AI has been used to support user interface and experience (UI/UX) designs targeting the cognitive search strategies and information processing limits of the humans using these systems. Although web search would not exist without AI, prior efforts that have used web search as a context for understanding and assessing digital literacies have primarily viewed search as the interaction between the individual and the source materials, ignoring the role of AI within search both in the assessment design and in what learners are expected to know and be able to do within search tasks.[44]

> However, having an understanding — or at minimum, an awareness — of the role of AI in presenting and summarizing source materials is an essential step required for independently and critically evaluating those materials.

While web search has undergone significant changes over the past 30 years,[45] the development of GenAI and LLMs stands to further reshape how AI technology is used to locate, access, evaluate and synthesize information.

As such, KSAs specific to the use of AI are becoming increasingly crucial in supporting individuals' ability to navigate and utilize these evolving digital environments critically and effectively.

As an illustrative example, we consider the use of web search to answer the inquiry question "Why is palm oil bad for the environment." We chose this question from a specially crafted "Researchy Questions" dataset, which includes multi-perspective search engine queries that are challenging for both humans and AI systems to answer using content on the web.[46] This inquiry question could be pursued via typical web search, or by prompting an AI-enabled search engine (see Figure 3) and thus offers a useful scenario in which to explore the potential to integrate AI literacy competencies into an assessment of digital literacies.

Figure 3 provides example screenshots of two different semantic and keyword-based search engines (Microsoft Bing and DuckDuckGo) alongside the responses from two different LLMs designed to support web search (Microsoft's Bing Copilot and OpenAI's GPT-3.5 Turbo). The contrast between the traditional search engine results page (SERP) from Microsoft's Bing engine and the privacy-focused search engine DuckDuckGo exemplify how the way in which search engines differ in how they integrate AI into the design of search result pages (left panels). For example, Bing places text and source summaries at the top of the SERP and recommends related searches. In contrast, the DuckDuckGo SERP makes less use of summarization and recommendation algorithms, using its own algorithms to retrieve and rank the sources.[47]

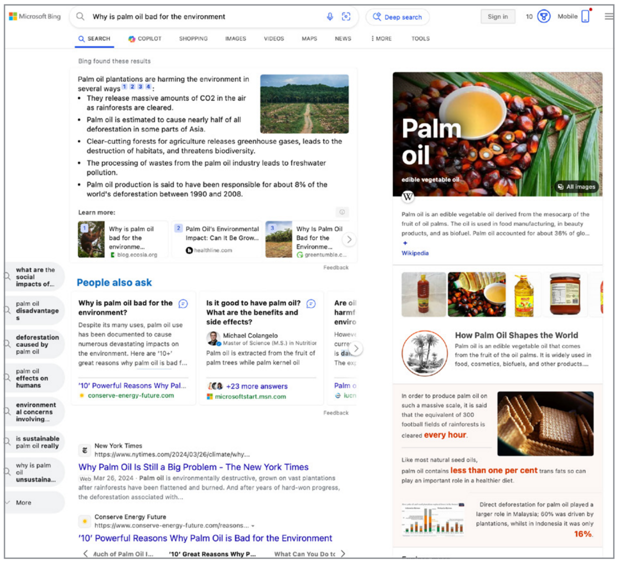
This contrasts with the two examples of conversational LLM-based search supported by Bing Copilot and OpenAI's GPT-3.5 Turbo hosted by DuckDuckGo (right panels).

While the content of the responses is similar, the Copilot chat is longer and has been augmented with citations, images and recommended follow-up questions to deepen or expand the inquiry. These differences highlight the rapid, ongoing research and development efforts focused on improving how LLMs can be used to support web search.[48]

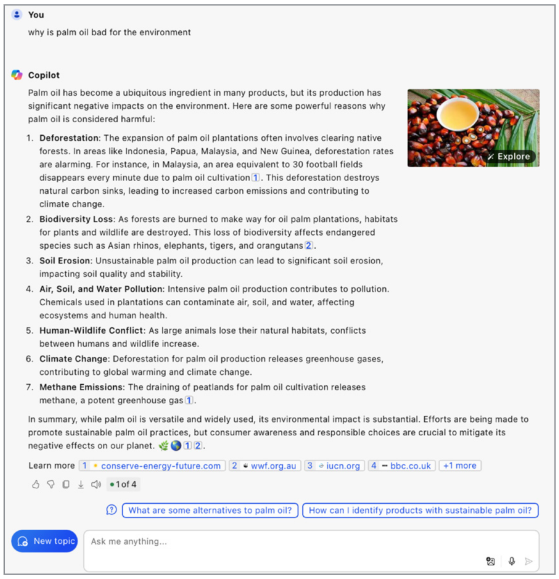
While we note that both the traditional and LLM-enabled search engines described here make use of AI, a key distinction is that the former primarily uses AI for information retrieval (in ways that are typically backgrounded for end users) and renders a list of direct links and source descriptions, while the latter involves direct user interaction with LLMs to generate a condensed summary across multiple sources (which may or may not be directly visible to users).

Figure 3. Example illustrating "Why is palm oil bad for the environment" query in traditional (left) and conversational (right) search conducted on Bing (top) and DuckDuckGo (bottom) on May 8, 2024.

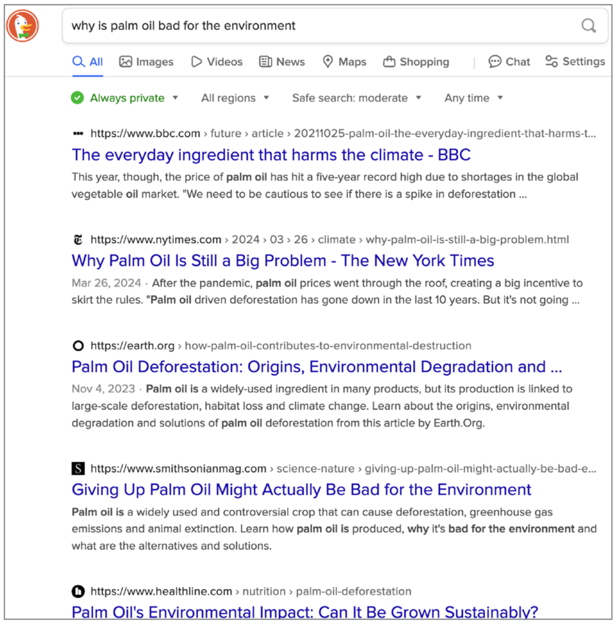
Traditional search results:
Microsoft Bing



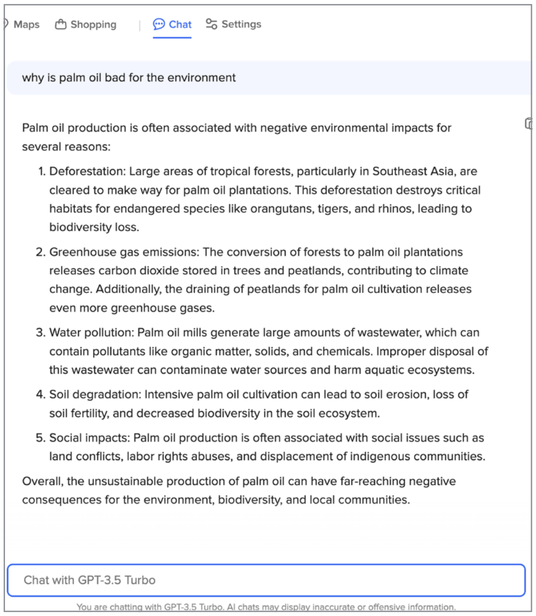
Conversational AI search results:
Microsoft Copilot



Traditional search results:
DuckDuckGo



Conversational AI search results:
GPT-3.5 Turbo



We provide these four examples not only to illustrate the diverse uses of AI to support and augment web search but also to motivate the consideration of AI literacy as it relates to a set of generalizable skills that can be considered as targets for digital literacies assessment across various contexts, as opposed to use of a particular tool reflecting a specific set of technology design decisions.

Considering the affordances AI provides within this web-based inquiry scenario, we next discuss how AI literacy skills may complement digital literacies in terms of (1) understanding the search task, (2) assessing relevance and quality of results and selectively accessing content, and (3) information synthesis.

Understanding the search task

The inquiry question "Why is palm oil bad for the environment" has a hierarchical structure, meaning that to answer it one must recognize and answer several other questions such as how palm oil is produced, the scale of the production and the effects of these practices.[49]

In Figure 3, the traditional Bing search with augmentation algorithms identifies some relevant questions that would enhance an understanding of this search task about the societal impacts of palm oil and deforestation; however, we also see many questions and suggested prompts related to health effects of palm oil that distract from the main goal of the inquiry. In contrast, LLM-based searches provide results as a conversational response, creating a shared context of information that can be further expanded to meet the individual's needs.

> Research comparing traditional and LLM-based tools found that individuals using LLM-based search were able to complete search tasks in almost half the time and with significantly more complex search queries.[50]

Though LLM-based search may be time efficient, there remains a risk that the resulting information may contain misleading information or confabulations, which users may overly rely on when they are not otherwise cued to potential inaccuracies in the LLM response.[51]

In both cases, users must determine if the information meets the needs of their inquiry task, though their approach to refining their inquiry may look quite different given the varying supports provided in the environment (e.g., using recommendations or further prompting).

AI literacy skill: Use and apply AI

Whether through complex prompting, building a shared context through conversation or the use of recommendation tools, AI can be used to modify or refine a search to enrich both the information retrieved and one's own understanding of the search task. Ultimately, however, it is up to the individual to recognize and use these tools to build their understanding of the information space.

Research exploring confirmation bias within web search found that individuals tended to express personal biases in their prompts for LLM-based searches, which in turn produced more polarized information, creating, as the authors put it, a generative echo chamber.[52] Assessments could measure the extent to which an individual's exploration of the information space reflects the exploration and consideration of multiple perspectives.

Assessing the relevance and quality of results and selectively accessing content

Existing digital literacies frameworks have primarily focused on identifying and evaluating sources of information as opposed to scrutiny of the algorithms that retrieve and present such information. In Figure 3, we see a stark contrast between the information presented by the two traditional search engines, with Bing presenting more information that is not directly relevant to the question (such as images of palm oil, the Wikipedia widget and excerpts from Q&A forums) while DuckDuckGo presents a simple list of sources with text previews akin to a more traditional search engine.

While both sites require the ability to assess relevance and quality, they vary in the type and number of cues available to inform that decision. LLM-based search further complicates this skill by synthesizing information from across the Internet in a way that obscures the original source(s), making it difficult to assess quality without further search. Specifically, LLM-based tools synthesize information not by accessing and searching the web directly, but by drawing on a pre-established knowledge base acquired during model training.[53]

AI literacy skill: Recognize, understand and evaluate uses and outputs of AI

The ability to recognize the use of AI in the retrieval and presentation of information and to evaluate whether the capabilities of the algorithm are aligned with the user's goals and expectations is becoming increasingly important as users are given more choice in how they retrieve information.[54] In an experiment comparing search tools, individuals using LLMs made worse decisions (i.e., choices that did not meet all task requirements), but when confidence estimates were assigned to statements provided by the LLMs (e.g., specific statements were highlighted with green or red to indicate high and low confidence information) the decision quality increased.[55]

It was hypothesized that confidence estimates supported users' general decision-making strategies of selectively attending to the most accurate information. Such AI literacy skills are complementary to existing digital literacies supporting information evaluation; however, there are still many questions surrounding how these skills might map to behaviors within an assessment task.

For example, how should an understanding of the risks of LLM hallucinations of source content[56] alter how we verify and use information generated by an LLM-based search tool? Likewise, how should user awareness of the susceptibility of search engines to search engine optimization (SEO) spam and other manipulations that may dilute the quality and relevance of results[57] affect how information is queried?

A related inquiry skill is that of sourcing, or the process of identifying and evaluating the source of the information and using that to evaluate the information itself. For example, in crowded and contentious spaces such as discussions of pandemics, health risks or environmental issues, learners must evaluate which information sources are reliable and trustworthy.[58]

Alas, most current LLMs do not reveal their sources, often because there is no direct mapping of responses to single origins.

Information synthesis

A digital literacies perspective on web-based inquiry tasks[59] emphasizes the importance of synthesis throughout the search process. Such synthesis supports deeper comprehension and learning.[60] Traditional web search is an iterative process where multiple documents are viewed and compared and the individual constructs a mental model reflecting the relationship among sources and an understanding of the information task.[61] In contrast, LLM-based conversational search tools appear to remove the need for multi-document synthesis by providing answers that seem to synthesize material from across the web. This is a case where appearances do not reflect reality[62] and can lead individuals toward making misinformed[63] or biased decisions.[64]

AI literacy skill: Navigate AI ethically

The synthesis of information via LLMs presents a complicated ethical challenge that extends beyond merely claiming an LLM-generated synthesis as one's own. Evolving research on the effects of using LLMs to support writing suggests using these tools can shape the opinions that users express and ultimately believe,[65] phenomena termed *latent persuasion*.[66]

As we consider AI literacy skills related to ethical use of AI, we expect that an awareness of the issues of bias, misinformation and potential for limiting diversity of thought and perspectives will be crucial to inform how these tools are used to support information synthesis. Issues around ethical access, plagiarism and source attribution – so central to digital literacies – also remain critical in the context of using AI tools effectively and appropriately to support synthesis.

SECTION 03.

Guidance for developing assessments of digital and AI literacies

Given the conceptualizations of digital and AI literacies and the illustration of how such skills may be integrated within digital inquiry tasks presented in previous sections, *how then should such skills be assessed in a way that ensures useful insights*?

To be useful, assessment designs should be relevant to and reflective of contemporary realities of innovation in ways that leverage affordances of new technologies within emergent sociotechnical contexts.[67]

It is essential to use principled design approaches and to carefully consider the contexts in which the assessment will be administered and the implications of those contexts for assessment design and development.

Principled design approaches

Principled approaches such as evidence-centered design (ECD)[68] should be employed to ensure that assessments meet standards of technical quality (validity, reliability and fairness).[69] ECD is a particularly useful framework for designing digital assessments that may incorporate multiple different sources of data about various KSAs enacted in complex digital environments because it encourages detailed specification of how interactions in the assessment environment may be linked to the larger claims we wish to make about learners through a process of reasoning about evidence collected via the assessment.[70]

Extensions of ECD have been developed to reflect the important role of context from a sociocognitive, situated perspective,[71] making this approach especially appropriate for application to the design of assessments of digital and AI literacies, which we have argued are inherently contextual.

Developing a conceptual assessment framework within an ECD approach involves specification of the *student model* or competency model that specifies the KSAs most relevant for the target construct(s), the *task model* that specifies what kinds of tasks and environments constitute appropriate contexts in which to elicit or demonstrate proficiency with those KSAs and the *evidence model* that articulates how observations from that environment can be evaluated to draw inferences and support claims about learners' proficiency with target constructs given their behaviors and responses within the assessment task.[72]

To effectively communicate about learners' performances, we must also carefully implement approaches for *score reporting* that contextualize assessment results and provide actionable feedback. We briefly discuss each of these components and exemplify how they may be implemented in an assessment of digital and AI literacies using the illustrative example of conducting LLM-enabled inquiry (Figure 3).

Student model: What digital and AI literacies should we measure?

A student model or competency model specifies the variables most relevant to the KSAs we wish to measure in the assessment.

What we measure in the student model is directly linked to the claims we wish to make about learners based on the assessment results. Considering the illustrative example, we may wish to make meaningful claims about how successful learners are in effectively and ethically interacting with and using AI-enabled tools in the service of conducting online inquiry within web search contexts (see the section on score reporting).

Accordingly, the student model may specify variables related to the KSAs required to perform in such contexts (Figure 2). In the example of LLM-enabled web search (Figure 3), this may include understanding the search task, creating or revising effective LLM prompts, assessing relevance and accuracy of search results, critically evaluating output from LLMs (and rejecting misleading or incorrect information) and synthesizing useful information obtained through LLM-supported search while avoiding plagiarism.

These student model variables may be organized within a hierarchical structure or ontology that enables specifying nested relationships among the different components.[73] Different tasks may require different combinations of KSAs or may require additional variables to be specified. Such variables allow us to characterize learners' KSAs to inform claims about proficiency levels for the various target competencies in the student model.

In defining a student model for digital and AI literacies, assessment developers must specify the claims they wish to make about learners and the corresponding KSAs that make up the relevant domain.

What should be measured in digital literacies assessments? Whether there is emphasis on promoting mastery or proficiency, interpretation or evaluation and critique, knowledge or demonstration of skills and practices in various contexts[74] largely depends on the purpose of the assessment, which is closely linked to the desired claims.[75]

For example, to support claims about individuals' ability to use AI tools to communicate novel information, assessments could engage learners in creating, producing and sharing knowledge with AI tools in social media contexts, rather than simply replicating or consuming information produced by AI.[76] The process of defining a student model includes answering two key types of questions: What are we assessing? (i.e., what does it mean to be AI literate?) and for what purpose are we assessing? (What do we want to be able to say about learners? Who are the learners? How will these inferences be used?)

To the extent that claims are to be made about groups of learners, the student model must be specified in such a way that it reflects a construct definition that is sensible and appropriate for those groups. In the context of large-scale international assessment, for example, it may be challenging to develop a student model that appropriately reflects digital and AI literacies constructs in ways that are generalizable across national contexts, given apparent differences in emphasis and conceptualizations across cultural and language groups.[77]

Despite this challenge, several international frameworks for digital literacies skills have been developed by organizations like OECD[78] and others.[79] Such frameworks have been successfully used to develop international assessments designed to compare levels of proficiency with target skills across contexts[80] (for example, they specify competencies at a general level and definitions are not aligned to a specific national curriculum or instructional approach). These frameworks could inform student models of digital and AI literacies that may generalize across contexts and that are well-aligned to the kinds of comparative claims that international assessments are intended to support.

Developmental factors that may influence construct definition and emphasis on particular digital and AI literacies KSAs also merit consideration.

Just as there have been prior efforts to characterize digital literacies competencies in terms of developmental trajectories of learning progressions,[81] there have also been efforts to define key features of AI competencies spanning from early childhood into adulthood[82] and to develop curricula on foundational AI concepts for young children.[83]

For example, during elementary school, emphasis could be placed on exploring and building awareness of AI topics, such as by interacting with simple toys or machines that operate based on principles of AI.

During middle school, emphasis could shift to gaining familiarity while more systematically and independently experimenting with increasingly more abstract conceptual aspects of topics related to AI by using AI tools and weighing in on the advantages and disadvantages of AI use in certain contexts.

In high school, fostering foundational knowledge of the technical or theoretical aspects of AI use in a variety of settings, gaining familiarity with more advanced technical and philosophical implications of AI use, and seeking to independently acquire and apply new knowledge related to AI could be emphasized.

Adult learners might therefore be able to apply abstract problem-solving involving AI and advance a theoretical understanding of AI tools, uses and implications. Ng et al.[84] also identified competencies associated with experiences, foundational knowledge and understanding societal impacts and ethics of AI across elementary, middle and high school levels. Such frameworks can help to specify developmentally appropriate student models.

Task model: Where should we measure digital and AI literacies?

A task model specifies features of tasks and contexts within which appropriate evidence of target KSAs can be elicited.

Task models provide guidance that enables assessment developers to design, author and implement tasks that provide appropriate opportunities to demonstrate those targeted competencies. Task model variables reflect required and optional features of activities and environments in which learners will interact and provide responses (e.g., characteristics of texts, images, tools, possible interactions, directions and item formats).

Task models also specify the form(s) learners' interactions or responses may take (e.g., in-task actions, selected responses, constructed responses, performance tasks, etc.). In the LLM-enabled search example, potential task model variables would include topics and inquiry questions; features of the search engine, SERP or LLM interface; characteristics of source materials in terms of relevance, credibility or other parameters; vocabulary, text length and text complexity and so on.

Each task model variable can have certain values (e.g., presence/absence, degree or quality of a given feature) enabling characterization of tasks in terms of the constellation of features that enable certain evidence to be collected to support desired claims. This flexibility enables multiple and varied tasks, or families of related tasks, to be created by instantiating different values of the task-model variables for each task.

Potential work products captured in the LLM-enabled search scenario would include prompts/queries entered, navigation actions taken (e.g., clicking links, revising prompts, using help menus), or selections on multiple-choice items that could be incorporated before or after an interactive search.

Thus, this task model would delineate *where* the skills of task understanding, searching and accessing sources, analyzing and evaluating LLM responses and synthesis of results would be measured, as well as *how* they would be captured in terms of learners' responses and behaviors.[85]

How, then, should tasks measuring digital and AI literacies be designed? We suggest that assessment designers consider using scenario-based tasks (SBTs) when developing task models for digital and AI literacies assessments. Scenario-based assessments[86] situate learners in authentic task contexts and rich scenarios that are reflective of the real-world practices about which we wish to make claims.[87]

When designed carefully (both at the backend and front end), using SBTs that integrate selected response questions (e.g., multiple choice, drag-and-drop) and performance-based activities (e.g., interacting with search engine or LLM prompt interfaces to access and evaluate information) can potentially help to tease apart complex KSAs and enhance the accuracy in assessing learners' proficiency with digital literacies.[88] SBTs offer a compelling interplay between process and outcomes.

SBT approaches are readily expanded to incorporate interactions with GenAI and LLMs to find, analyze, evaluate and synthesize information as in the illustrative example (Figure 3). Specifically, learners' use of LLMs can be viewed as a sequence of choices that they make.[89] Thus, effective task models should define tools and task features that support interactions of interest (e.g., allow choice in how to interact with or navigate within search tools).[90]

For instance, the illustrative example could form the basis of an SBT in which learners are initially asked to answer selected response questions about their knowledge of how LLMs and search engines work. Next, they can be presented with an interactive inquiry task that asks them to perform a simulated web search (including using LLM prompting) to investigate the reasons why palm oil is bad for the environment however they see fit within constraints of the search tool.

After finalizing their search, they could be presented with questions that elicit further evidence about their search behaviors and their ability to synthesize conclusions from the results obtained (e.g., while avoiding plagiarism or direct replication of LLM output).

Notably, there are three distinct options for incorporating LLMs in the task model. At the most basic level, one could design scripted, discrete tasks that imitate LLMs. At an intermediate level, learners could directly interact with and use AI tools to respond to challenges given to them. At the most elaborate level, an LLM could be used to author tasks (based on a template) and evaluate learners' responses to it.

In all cases, the task model may be similar in terms of the challenge given to learners and design space and options that learners have. However, they reflect a key tradeoff between certainty (controlled by design) and authenticity (with a high degree of learning autonomy and potential unpredictability).

While avoiding the use of LLMs as an assessment tool streamlines the assessment process, it may become too synthetic and rigid as these technologies continue to evolve. On the other hand, tasks authored by LLMs can offer personalized and context-sensitive assessments but may support less accurate inferences. Use of LLMs to create conversation-based assessments following ECD principles is being explored.[91]

To make valid inferences about learners' KSAs and behaviors when completing performance-based digital tasks, it is important to design those tasks with process data modeling and analysis in mind.[92]

Process data refers to the extracted process indicators from the detailed logs of learners' interactions with the digital assessment platform, ideally capturing every state change including test-taker interaction together with their timestamps.

Developing the task model with process data modeling and analysis in mind includes the following iterative steps:[93]

- model construct-relevant mental processes needed to perform the task (e.g., encoding instructions, retrieving relevant information from memory, performing the necessary problem-solving or reasoning steps, responding)[94]
- embody theorized construct-relevant problem-solving or reasoning steps in external actions as much as possible and make sure that those actions and system changes are logged properly at the backend
- generate and document research questions and hypotheses related to expected action sequences for different proficiency levels
- conduct think-aloud, usability and/or pilot studies to test the alignment between logged interactions and observed behaviors
- model and analyze process data in line with the hypotheses, use additional data-driven approaches to make further construct-related discoveries and report the confirmatory and exploratory findings at a granular level[95]

In the LLM-enabled search scenario, construct-relevant mental processes may include:

- encode the task instruction and select a strategy from their repertoire to provide input to the search engine
- (re)type the input
- submit the search (click or hit enter)
- encode search engine results
- retrieve knowledge from declarative memory indicating that LLM-based summary or search results might be incorrect or biased
- apply visual search strategies to find reliable sources among the individual results or use the traditional web search to evaluate the LLM-based web search (if the task also includes traditional search tools)
- click the individual results and encode their content to gather evidence
- accumulate evidence by reasoning about and synthesizing the results
- refine the search input if more evidence is needed

Designing with process data modeling and analysis in mind also includes creating front-end interactions that can be logged at the backend; these front-end features must also be specified in the task model.

For example, presenting the task instruction on a separate screen than the web search task allows us to infer when and how long learners go back to the task instruction to remind themselves about the task specifics; logging each keystroke allows us to make inferences about search input strategies at a granular level (e.g., revision before hitting enter); logging scrolling depth allows us to make inferences about what part of the search results learners are evaluating; logging individual clicks on the results and presenting the results on a separate page allows us to monitor their evidence-gathering processes.

Presenting the final response on a separate screen helps us to infer the timing and the context for when learners think that they gathered enough evidence or realized they did not have enough evidence and return to the search screen.

Before leveraging process data for scoring or reporting purposes, think-aloud, usability and/or pilot studies can be conducted to ensure that learners interact with the task as expected and necessary interactions are logged.

Evidence model: How do we measure digital and AI literacies?

An evidence model provides specifications for how to interpret the work products and observable behaviors created by learners as they engage in an assessment to inform inferences about variables included in the student model (e.g., levels of a particular KSA given a particular response or observed behavior).

Evidence models consist of two components: a set of *evidence rules* or *scoring rules* that enable evaluation of observable work products in terms of the student model variables targeted in a particular task and a *measurement model* (i.e., statistical and psychometric models) that supports probabilistic inferences about learners' KSAs given the observed performance by linking student model variables to observables.[96]

In the context of LLM-enabled web search (Figure 3), an evidence model would specify rules (e.g., scoring rubrics, scoring algorithms) for how to evaluate learners' search strings, LLM prompts and iterative revisions to searches/prompts in light of target competencies (e.g., use and apply AI, refine prompts/queries as needed).[97]

The measurement model would then specify statistical approaches (e.g., using item response theory)[98] to link observables back to the student model (i.e., updating student model variables given the observations) and for accumulating evidence across observations to create probabilistic estimates of learner ability. While it is beyond the scope of this paper to provide detailed discussion of measurement models, we offer perspective on potential approaches to defining scoring rules for digital and AI literacies assessments given the competencies and tasks discussed previously.

Clear and comprehensive scoring rules are a crucial component of designing and implementing digital assessment tasks to measure complex KSAs like digital and AI literacy. Scoring rules can vary from detailed rubrics or coding schemes to simple rules based on correctness (e.g., 1 = correct, 0 = incorrect).

These rules can address both learners' processes (i.e., actions and action sequences) and work products (i.e., outcomes or item responses). Scoring response products focuses on the final work product(s) a learner produces and/or their responses to selected response or open-ended questions.

For example, a selected response question could be scored as full credit if a learner selected the correct response (e.g., = 2), partial scores given if a learner selected a partially correct response (e.g., = 1) and no credit given if a learner gives an incorrect response (e.g., = 0). Scoring process data requires evaluating learners' action steps and interactions during the task.

This involves capturing and analyzing interactions over time, such as the sequence of actions, the frequency of tool use and the time spent on various parts of the task. These process-based indicators can support inferences about learners' problem-solving strategies and behaviors and can be incorporated into final scores (e.g., iSkills assessments[99]).

In the LLM-enabled search example (Figure 3), an action sequence including interacting with the relevant individual search results to gather evidence, resubmitting a search when there is not enough evidence and finally coming up with a response could be scored as full credit, while an action sequence indicating that a learner submitted a search and then copied and pasted the LLM-generated output as a final response could receive no credit.

Before defining scoring rules for process-based indicators, it is important to validate the expected behaviors with think-aloud and interview studies with learners from the target population. Importantly, if the process data will be incorporated into scoring, learners should be notified that their interactions will be evaluated as well as the outcomes. This transparency helps to ensure the validity of the interpretations of the process data and scores.

To support claims made about learners' abilities, it is also necessary to identify and aggregate various pieces of evidence. This can include aggregating different process data indicators (e.g., sequences of construct-relevant actions learners take together with their timing) and/or process data indicators with product data (i.e., answer to the response). As SBTs are highly contextual, it is important to use several layers of triangulation:[100]

- *Different indicators for similar skills*: To ensure that interpretation does not rely on single operationalization of constructs, multiple indicators for each construct should be used.
- *Similar indicators across tasks*: As tasks are rooted in domains, prior knowledge and situational interest impact engagement and learning. Thus, indicators from different domains should be examined.
- *Process and outcome level indicators*: A complementary approach supports understanding of the relationship between different process patterns and their impact on outcomes and overall ability.

Overall, incorporating these elements into clear and comprehensive scoring rules, together with tasks that are designed with process data modeling and analysis in mind, can ensure that assessments of digital AI literacies provide an evaluation of learners' KSAs that assess not only the final outcomes but also the processes that learners use to arrive at the final product, offering a deeper insights into what they know and can do with AI in digital literacy contexts.

Score reporting: How to support interpretations of performance

A related but distinct issue from scoring rules concerns how to present and communicate assessment results to learners and other users of test scores. Score reports should clearly present assessment information to learners, teachers and other audiences and should support appropriate decision-making. Assessment information may include total score and subscores for the construct(s) being assessed, information to support interpretation of these scores and potential next activities given the results.

Depending on the type of assessment and the evidence collected about learners' KSAs, different types of claims can be produced.[101] For digital and AI literacies, well-designed SBTs may provide enough evidence to produce scores for an overall digital literacy construct and its subconstructs (e.g., critical evaluation of digital resources in the context of AI or creating information products in the context of AI).

Contextual information generated by the assessment scenarios can be used to provide additional information to help individuals understand the meaning of the scores. This may include using process data to provide insights on how learners responded to the tasks, strategies used, timing data (i.e., as a proxy for engagement[102]) and response patterns associated with certain misconceptions. Performance could also be compared against emergent learner profiles or common patterns observed (e.g., learners who took a similar approach to the task also tended to demonstrate certain KSAs), enabling more personalized, rather than standardized, score reporting.

An interesting application of LLMs is to support the implementation of personalized score reports for different audiences.[103] Using LLMs, it is possible to generate explanations (i.e., narratives) describing the main highlights of the score report or an interactive dashboard, as well as potential misinterpretations, to increase the explainability of the findings.[104]

This type of approach could also support more actionable reporting to make assessments of digital and AI literacies more useful for individuals who take them. Further, we recommend that when providing individual-level score reports, institutions seek reports that include performance feedback: descriptive, qualitative information about learners' performance on digital literacies tasks, rather than atomistic, quantitative information, to help score users better understand learners' strengths and weaknesses for improvement purposes.

For example, contextualized score reports for the illustrative example (Figure 3) might provide information on whether a final product includes only accurate information (or includes LLM-based hallucinations), offers a complete answer to the inquiry task and is well-organized and communicative. This information can provide insights on potential next steps or actions by individuals at different levels that could result in improving learners' AI literacy levels (e.g., practice strategies for verifying accuracy of LLM-based information, monitoring work products to ensure all task demands are met or improving organization and development in writing).

Contextual considerations for assessment

Several notable efforts have made progress toward identifying competencies and curriculum for digital and AI literacies across the globe. For example, Ng et al.[105] identified initiatives in Europe (Erasmus AI+), Hong Kong (AI for the Future), Singapore (AI Singapore) and the U.S. (Digital Promise and AIK12). When identifying critical competencies for international contexts, there are several important challenges to consider.

As previously mentioned (see Student model), there have been prior efforts to identify cross-cultural core competencies[106] as a critical step in this direction. Cross-cultural foundational skills for AI literacy within the context of digital literacies likely include: foundational understanding of the AI technologies; how to use, create and share information using AI technology and being able to consider the ethical and social impact of using such technologies.[107]

Beyond defining generalizable competencies, it is essential to recognize how contextual factors may shape assessment of these literacies.[108] These factors include access to technology, linguistic and cultural aspects of learning, local educational practices and systems, policy and governance, as well as economic factors.

Access to technology

The so-called "digital divide"[109] both between and within countries, affects access to technology and the Internet, creating unequal opportunities for learning digital skills. To bridge disparities in access to digital technologies, it is essential to develop assessment frameworks that can be used toward identifying, acknowledging and addressing different levels of access.

For example, providing offline alternatives and low-cost solutions may be appropriate in settings where disparities in digital access have been documented. To ensure assessment frameworks and construct definitions are equitable, it is worth considering whether evidence of digital literacy will appear comparable across both high- and low-technology resourced areas.

Linguistic diversity

A primary challenge in assessing complex constructs such as the various digital literacies previously mentioned across different cultures stems from the variation in the language itself. Certain terms and concepts in digital and AI literacies may not translate well if treated in a literal sense.[110] To address this, it is crucial to develop a conceptualization and framework for cross-cultural assessment in close partnership with content and language experts from tested groups to ensure the appropriate translation of key terms to preserve their intended meaning.

Cultural attitudes, norms, values and relevance of content

Setting linguistic diversity aside, another challenge is that educational content may not be universally applicable or relevant across different cultural contexts, even within the same country. Cultures may place distinct emphasis on certain aspects of digital and AI literacies that are perceived as important in their context.[111]

Cultural attitudes toward technology, privacy and data security also could vary widely given they are influenced by factors such as trust in technology and societal values. What is considered normative will vary across cultural groups.

Local educational practices and systems

Teaching methods, pedagogical practices and instructional standards may also differ widely across countries and cultures. Additionally, educational systems and infrastructure vary significantly in terms of organizational structures, curricular emphasis and types of resources available. Therefore, digital and AI literacies assessment frameworks must be adapted to align with local educational standards and practices.

A universal conceptualization of digital and AI literacies should be flexible enough to fit local teaching practices and systems, providing a roadmap on key digital skills that could be incorporated into an existing curriculum. Gaining support from all interested stakeholders is critical to be effective in this regard.

Policy and governance

National policies regarding digital education and AI literacy can differ,[112] impacting the implementation of literacy programs. Collaborating with local policymakers to ensure that digital and AI literacies frameworks align with national and international education goals and policies is essential. Such regulations have implications for the extent to which AI literacy must be measured, the types of data that can be collected from or used to evaluate user interactions with AI systems, the involvement of users in development, etc.[113]

Furthermore, political stability and governance impact the adoption and integration of digital literacy programs. Developing flexible conceptualizations of literacies that can be adjusted to varying sociopolitical contexts can ensure continuous learning opportunities within and across borders.

Economic factors

The emphasis on certain literacy skills within digital contexts may vary depending on more distal factors, such as which skills are most likely to be in high demand within the local or regional workforce. While there is general concern about the impact of AI-driven automation across industry sectors and regions, the extent to which certain job categories may be impacted is likely to vary across contexts.[114] This has obvious implications for identifying the knowledge and skills most likely to prepare learners to be successful in future occupations and professions.

SECTION 04.

Risks and consequences of digital and AI literacies assessment

Our discussion has largely focused on the need to assess competencies related to the use of digital media and technologies, especially those involving interactions with AI-based tools. Identifying which competencies individuals have acquired can help inform strategies to support them in addressing potential knowledge and skill gaps, such as instructional interventions or policies supporting skill development.

Ideally, these insights from assessments could be used to support individuals in gaining the knowledge and skills needed to thrive in the current and future workforce and to inform data-driven policies that effectively close gaps created by systemic inequities by ensuring that learning opportunities are available to all.

While these intended consequences reflect an ideal scenario, there is a risk that defining and assessing a set of complex competencies at scale can have unintended negative consequences (see Table 2). To mitigate potential negative consequences, it is therefore important to identify potential risks associated with assessing literacy competencies that involve the use of digital media and technologies, particularly those involving AI.

In this section, we identify several risks and questions to consider in developing assessment approaches to measure relevant digital and AI literacies.

TABLE 2. Risks of digital and AI literacies assessment

Narrow focus	Risk of defining competencies in a way that does not adequately capture crucial relationships to other foundational literacy skills.
Inequity	Risk of developing frameworks that widen or introduce inequity.
Instability	Risk of AI systems providing inconsistent answers over time, making scoring unreliable.
Lack of creativity	Risk of diminishing opportunities for learners to develop or demonstrate creative, divergent thinking.
Limited generalizability	Risk of designing frameworks that cannot be adapted in global contexts.

First, there is the risk of defining or operationalizing the focal construct(s) in a manner that does not adequately capture the full range of competencies in a way that supports an evidence-based approach to assessment (i.e., construct underrepresentation as a threat to assessment validity).[115] This underrepresentation could create challenges in identifying appropriate and valid sources of evidence of competencies that ideally uniquely tap into the target construct. Under the ECD approach, an ill-defined construct threatens the integrity of the student model and ultimately the foundation of the assessment.

A useful analogy may be found in the emergence and widespread use of automotive technologies over a century ago: learners need to know basic principles of how cars work in a way that helps them identify where and when it should be used (understanding AI), how to drive the car within a set of given ethical and legal constraints (using AI), and more broadly, thinking about the impact of cars on society, identifying disparities in access to cars and possible solutions for further innovating automotive technologies in a way that reduces any harmful consequences of car use (impact of AI).

Requirements for a drivers' license may emphasize measures of understanding (written test) and use (driving test) versus larger societal impacts; however, in the case of AI literacy, all three aspects of the construct may be important to develop and to assess. This is a unique challenge, particularly across contexts and as AI-driven digital technologies continuously evolve.

AI literacy is a "moving construct" in that focal competencies today may not be so important in a few years; as technologies evolve, so too must the literacies required to effectively use those technologies to understand, create and communicate. This is particularly concerning given the evidence that educators tend to adapt instruction around what content is tested (i.e., teaching to the test).[116]

Further questions to consider include:

- What knowledge of AI is required to effectively interact with a world that's heavily digitally mediated (e.g., knowledge of technical aspects of AI vs. knowledge of how to navigate a task using AI-driven inputs and outputs)?[117]
- What skills and knowledge do we expect learners to have when interacting with AI?
- How do we distinguish between what skills a learner has and what skills they have yet to develop, now that AI can be used to automate or simplify some target skills?

Related to the issue of defining the construct, it is important to recognize that an individual's existing KSAs and experiences affect how they interact with new technologies. Given the overlap noted with traditional literacies, individuals with lower levels of core literacy skills may not benefit as much from the use of AI technologies if left unsupported.[118]

Consideration of what foundational literacy competencies (e.g., decoding), background knowledge, critical thinking and evaluation skills and technical knowledge about AI that are necessary to grasp essential digital and AI literacies concepts is crucial for developing assessments that produce fair, valid and interpretable results.

We may also consider questions such as the following:

- What relationships do digital and AI literacies have with other related constructs (e.g., between AI literacy and reading literacy)?
- To what extent is it critical to consider foundational reading skills in gauging digital and AI literacies?
- Is it still necessary to measure foundational writing and synthesis skills?
- Should we adapt our expectations for foundational literacies to align to new tools and contexts (e.g., skills for crafting good AI prompts as opposed to writing essays)?
- If a greater emphasis is placed on literacies involving AI technologies, what other key competencies might be subsequently de-emphasized or left out of the curriculum?

Second, there is the risk of developing assessment approaches that might effectively widen or introduce additional sources of inequity, rather than provide information that can be used to document or mitigate existing inequities (see Table 3). Many of the factors that contribute to the so-called "digital divide" (e.g., learners' motivation to use technologies, technology materials and resources available, technical skills, prior usage and familiarity)[119] are important to consider in developing assessments of AI literacy competencies in digital contexts. How they are considered likely depends on the purpose of the assessment.

For example, those who can successfully integrate AI-based tools into their learning and work are more likely to excel relative to peers who have not had opportunities to develop foundational skills and therefore struggle to gain proficiency. On the other hand, an intended consequence of well-designed digital literacies assessments may be to provide learners an opportunity to develop skills such as the ability to evaluate and critique information produced by GenAI tools in digital contexts (i.e., assessment for learning).[120]

TABLE 3. Equity in digital and AI literacies

Digital divide	Unequal access to technology affects the ability to develop and assess digital and AI literacies.
Cultural norms and values	Assessments must reflect diverse perspectives, especially in large-scale international contexts.
Equity in education	Teaching and assessing digital and AI literacies is crucial for equity, reducing vulnerability to unethical technology use.

However, we cannot assume that all learners will equally benefit from such opportunities, particularly if the assessments are not instructionally aligned to an equal extent for all learners[121] or if learners lack sufficient foundational literacy skills or background knowledge to effectively engage with the higher level task.[122] Some questions to consider in this regard may include:

- How can assessments be designed to ensure equal access and opportunity to demonstrate KSAs for all test-takers, regardless of their technological resources?
- Can assessments be well-aligned with the instructional goals and outcomes to an equal extent across contexts and for all test-takers?
- How might the assessment address the relevant core competencies of test-takers with varying foundational literacy skills or background knowledge?

Third, while GenAI and LLMs are good "pattern detectors" in many instances, patterns may be complex and thus require a more nuanced, sensitive or creative interpretation. Furthermore, there is the possibility that such AI-driven tools will produce information that could be unstable and difficult to replicate (i.e., an LLM may give different answers to identical prompts at different times), which may pose challenges for applying consistent scoring rules across administrations of a given assessment.

Though responses generated by AI may appear to be coherent at the surface, they may include counterfactual or unverifiable information. For example, the specificity of information contained in the prompt may change the results in widely disparate ways and a nonexpert may need guidance in understanding how to effectively write prompts.[123] Though the "temperature" or stability of the output created using AI can be adjusted, individuals need to know that the feature exists, its purpose, how to control it and how to interpret the output considering its setting.[124]

Additional questions to consider include:

- How do we ensure that learners can develop durable skills like critical thinking while also knowing when and in what contexts it is appropriate to use/learn with AI?
- Considering content created through GenAI/LLMs may not be replicable, what skills should be assessed to ensure learners' ability to discern accurate information, while building a collective knowledge base through interactions with peers?
- How do we use AI to access and use information more efficiently, while simultaneously not depending on it? Similarly, how do we continue to be active processors, critical evaluators and co-constructors of information and while not surrendering our human agency and become passive recipients of information?

Fourth, yet another potential long-term unintended consequence of increasing use of AI-based tools to support digital literacies is a potential narrowing of ideas — essentially a reduction of divergent or creative thinking (e.g., groupthink) — due to the nature of how the iterative use of LLMs operates. Responses created through prompting LLMs reflect the underlying data on which such models are trained, which are likely to reflect a sort of "average" and more likely to rely on common or dominant perspectives.

Without knowledge of how to specifically prompt for alternative perspectives, or to apply critical thinking and evaluation, there is a potential risk that AI users over time will rely on and reinforce some common set of ideas that tend to appear in the training datasets, leading to diminished outputs[125] and potentially diminishing future innovations over time. There is also the risk of diminishing opportunities for learners to develop the skills that are likely to result in creative outputs. Questions for further consideration may include:

- How could assessment tasks encourage critical thinking and questioning of AI-generated responses?
- Are there specific tasks, questions or prompts that could be designed to challenge the "average" or common perspectives generated by LLMs?
- How could assessment tasks evaluate learners' skill in effectively prompting LLMs for diverse perspectives?
- In what ways could collaborative tasks be implemented to encourage creative or divergent thinking or to foster innovative task solutions?

Fifth, there is the risk that assessment approaches may not be designed in a manner that can be implemented or adapted in cross-cultural or global contexts. As noted, there are multiple factors to consider when identifying key features of target constructs, especially those as complex as digital and AI literacies, across many contexts that differ in several key dimensions (see Contextual considerations).

Assessments that are to be implemented in international or cross-cultural contexts should be designed in a way that is responsive to the distinctive contextual characteristics of each setting.[126] When the goal of such an assessment is to produce comparable scores between examinees in a manner that upholds high standards for validity fairness and equity, this can pose obvious challenges. Some questions to consider may include:

- Are there mechanisms for adapting the assessment to fit different contexts?
- What steps can be taken to ensure that the content is culturally relevant and not biased toward any particular group(s)?
- How could an assessment accommodate linguistic diversity? How could it still provide equal opportunities for non-native speakers to express their skills?
- What innovative statistical methods or procedures for standardization could be used to ensure flexible implementation as well as the comparability of scores?

Finally, there is the risk that assessing individuals on competencies that rely on technologies may encourage them to use such technologies in a manner that reveals currently unforeseen ethical and/or legal concerns. This is particularly concerning for assessments geared toward youth. Such risks include those with respect to privacy and data ownership. As with any technology, there may also be unforeseeable downstream risks that could even be contradictory to the proposed benefits of the technology.

For example, social media came into widespread use as a tool for developing and maintaining social connections; however, more recently greater use of social media among youth has been found to be a key contributor to their mental health burden.[127] It is also important to recognize that ethics and legal concerns tend to vary in nontrivial ways across international and cultural contexts, an issue that is further complicated by its direct connection to historical precedent and technological advancement generally outpacing legal codification. Relevant questions to consider include:

- What risk management strategies are included to address both current and potential future ethical and legal issues?
- How can assessment use be monitored so that negative impacts can readily be addressed?
- What strategies are in place to continually update the assessment to align with the latest technological and legal developments?

SECTION 05.

Conclusions

We encourage further discussion, debate and research on how to measure digital and AI competencies in ways that power human progress.

In this report, we have offered recommendations for how to best assess proficiency with AI literacy as conceptualized within the larger umbrella of digital literacies. While digital literacies competencies remain of crucial importance, emergent GenAI tools and technologies will change how those KSAs are defined and applied to solve information-based problems and to create and communicate using digital technologies.

There are inherent complexities in these constructs, introducing challenges and opportunities for assessment. Approaches to the assessment of these digital and AI literacies are informed by construct definitions, task designs and scoring rules specified by developers, which are driven by the intended purpose and use case for a given assessment.

Within contemporary globally networked society, where large-scale assessments may be used to make claims about proficiency levels of individuals or groups across states, nations and continents, it is also essential to consider the role of contextual factors in informing assessment design, in terms of the importance and relevance of GenAI and LLMs to support digital literacies in this context.

Other contexts (e.g., classroom formative assessment) will have other needs or considerations that affect constructs, tasks and evidence collected. Awareness of the consequences of developing large-scale assessments of digital and AI literacies is necessary to guide the process of developing fair and valid assessments with potential to positively impact policy and practice.

We hope these reflections inspire discussion, debate and further research and development concerning measurement of these crucial digital and AI competencies in ways that can enable greater human progress and potential to be realized by leveraging affordances of these emerging capabilities.

SECTION 06.

References

[1]OECD, "The Future of Artificial Intelligence," in *Digital Economy Outlook 2024*, vol. 1, *Embracing the Technology Frontier* (Paris: OECD Publishing, 2024), 39–66, https://doi.org/10.1787/a1689dc5-en.

[2]Andrew Green, *Artificial Intelligence and the Changing Demand for Skills in the Labour Market*, OECD Artificial Intelligence Papers No. 14, (Paris: OECD Publishing, 2024), https://doi.org/10.1787/88684e36-en.

[3]Paul Gilster, *Digital Literacy* (New York: Wiley & Sons, 1997).

[4]Julie Coiro, "Toward a Multifaceted Heuristic of Digital Reading to Inform Assessment, Research, Practice, and Policy," *Reading Research Quarterly* 56, no. 1 (2020): 9–31, https://doi.org/10.1002/rrq.302; Kewman M. Lee et al., "Multidimensional Approaches to Examining Digital Literacies in the Contemporary Global Society," *Media and Communication* 7 (2019): 36-46, https://doi.org/10.17645/mac.v7i2.1987.

[5]Elena Forzani and Jill Castek, "Expansive, Dynamic, and Critical New Literacies: Widening Views of Digital and Multimodal Learning," in *International Encyclopedia of Education*, 4th ed., eds Robert J Tierney, Fazal Rizvi and Kadriye Ercikan (Elsevier, 2023), 10:487–496, https://doi.org/10.1016/B978-0-12-818630-5.07024-X.

[6]Colin Lankshear and Michele Knobel, *New Literacies: Everyday Practices and Social Learning* (New York: McGraw-Hill, 2011); Lee et al. "Examining Digital Literacies"; Don Leu et al., "New Literacies: A Dual-Level Theory of the Changing Nature of Literacy, Instruction, and Assessment," in *Theoretical Models and Processes of Reading*, 6th ed., eds. Donna E. Alvermann, Norman J. Unrau, and Robert B. Ruddell (Newark, DE: International Reading Association, 2013), 1150–1181, https://doi.org/10.1598/0710.42.

[7]Ester van Laar et al., "The Relation between 21st-Century Skills and Digital Skills: A Systematic Literature Review," *Computers in Human Behavior* 72 (2017): 577–588, https://doi.org/10.1016/j.chb.2017.03.010.

[8]Julie Coiro, Jesse R. Sparks, and Jonna M. Kulikowich, "Assessing Online Collaborative Inquiry and Social Deliberation Skills as Learners Navigate Multiple Sources and Perspectives," in *Handbook of Multiple Source Use*, eds. Jason L. G. Braasch, Ivar Bräten, and Matthew T. McCrudden (New York: Routledge, 2018), 485–501; Jesse R. Sparks and Paul Deane, *Cognitively Based Assessment of Research and Inquiry Skills: Defining a Key Practice in the English Language Arts*, Research Report No. RR-15-35 (Princeton, NJ: ETS, 2015), https://doi.org/10.1002/ets2.12082; Jesse R. Sparks, Irvin R. Katz, and Penny B. O'Neil, *Assessing Digital Information Literacy in Higher Education: A Review of Existing Frameworks and Assessments with Recommendations for Next-Generation Assessment*, Research Report No. RR-16-32 (Princeton, NJ: ETS, 2016), https://doi.org/10.1002/ets2.12118.

[9]Sparks, Katz, and O'Neil, *Assessing Digital Information Literacy*, 3–6.

[10]See Sarit Barzilai and Yoram Eshet-Alkalai, "The Role of Epistemic Perspectives in Comprehension of Multiple Author Viewpoints," *Learning and Instruction* 36 (2015): 86–103, https://doi.org/10.1016/j.learninstruc.2014.12.003; Sarit Barzilai et al., "Making Sources Visible: Promoting Multiple Document Literacy with Digital Epistemic Scaffolds," *Computers & Education* 157 (2020): 1–17, https://doi.org/10.1016/j.compedu.2020.103980; Saskia Brand-Gruwel, Iwan Wopereis, and Yvonne Vermetten, "Information Problem Solving by Experts and Novices: Analysis of a Complex Cognitive Skill," *Computers in Human Behavior* 21, no. 3 (2005): 487–508, https://doi.org/10.1016/j.chb.2004.10.005; M. Anne Britt, Jean-François Rouet, and Amanda Durik, *Literacy beyond Text Comprehension: A Theory of Purposeful Reading* (New York: Routledge, 2018), https://doi.org/10.4324/9781315682860; Byeong-Young Cho, Peter Afflerbach, and Hyeju Han, "Strategic Processing in Accessing, Comprehending, and Using Multiple Sources Online," in *Handbook of Multiple Source Use*, eds. Jason L. G. Braasch, Ivar Bräten, and Matthew T. McCrudden (New York: Routledge, 2018), 133–150, https://doi.org/10.4324/9781315627496-8; Julie Coiro, "Predicting Reading Comprehension on the Internet: Contributions of Offline Reading Skills, Online Reading Skills, and Prior Knowledge," *Journal of Literacy Research* 43 (2011): 352–392, https://doi.org/10.1177/1086296X11421979; Coiro, "Toward a Multifaceted Heuristic"; Coiro, Sparks, and Kulikowich, "Assessing Online Collaborative Inquiry"; Susan R. Goldman et al., "Literacy in the Digital World: Comprehending and Learning from Multiple Sources," in *Bringing Reading Research to Life*, eds. Margaret G. McKeown and Linda Kucan (New York: Guilford, 2010), 257–284; Sparks and Deane, *Cognitively Based Assessment;* Caitlin Tenison and Jesse R. Sparks, "Combining Cognitive Theory and Data Driven Approaches to Examine Students' Search Strategies in Simulated Digital Environments," *Large-Scale Assessments in Education* 11 (2023): 1–39, https://doi.org/10.1186/s40536-023-00164-w.

[11]Ou Lydia Liu et al., "Navigating the Digital Horizon: A Proposed Framework for Defining and Measuring Digital Literacy" (working paper, ETS, Princeton, NJ, 2024).

[12]See Britt, Rouet, and Durik, *Literacy beyond Text Comprehension*; Coiro, "Toward a Multifaceted Heuristic"; Luci Pangrazio, Anna-Lena Godhe, and Alejo González López Ledesma, "What Is Digital Literacy? A Comparative Review of Publications across Three Language Contexts," *E-learning and Digital Media* 17, no. 6 (2020): 442–459, https://doi.org/10.1177/2042753020946291; John Sabatini, Tenaha O'Reilly, and Nancy A. Doorey, *Retooling Literacy Education for the 21st Century: Key Findings of the Reading for Understanding Initiative and their Implications* (Princeton, NJ: ETS, 2018).

[13]For example, Pangrazio, Godhe, and Ledesma, "What Is Digital Literacy?"; Julian Sefton-Green et al., *Establishing a Research Agenda for the Digital Literacy Practices of Young Children: A White Paper for COST Action IS1410* (Brussels: European Cooperation in Science and Technology, 2016), https://www.lse.ac.uk/media-and-communications/assets/documents/research/projects/p4df/COST-2016.pdf.

[14]Marko Grobelnik, Karine Perset, and Stuart Russell, "What Is AI? Can You Make a Clear Distinction between AI and non-AI systems?" OECD, AI Policy Observatory, March 6, 2024, https://oecd.ai/en/wonk/definition; see also European Union, European Commission, *The EU AI ACT*, adopted April 21, 2024, https://www.euaiact.com/.

[15]J. Michael Helm et al. "Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions," *Current Reviews in Musculoskeletal Medicine* 13 (2020): 69–7 https://doi.org/10.1007/s12178-020-09600-8.

[16]"Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations," U.S. Department of Education, Office of Educational Technology, accessed June 10, 2024, https://tech.ed.gov/ai/.

[17]Brian Kennedy, Alec Tyson, and Emily Saks, *Public Awareness of Artificial Intelligence in Everyday Activities*, Pew Research Center, February 15, 2023, https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities/.

[18]Pekka Mertala, Janne Fagerlund, and Oscar Calderon, "Finnish 5th and 6th Grade Students' Pre-Instructional Conceptions of Artificial Intelligence (AI) and their Implications for AI Literacy Education," *Computers and Education: Artificial Intelligence* 3 (2022): 1–7, https://doi.org/10.1016/j.caeai.2022.100095.

[19]Anna Milanez, *The Impact of AI on the Workplace: Evidence from OECD Case Studies of AI Implementation*, OECD Social, Employment and Migration Working Paper 289 (Paris: OECD Publishing, 2023), https://doi.org/10.1787/1815199X.

[20]See Michael J. Handel, "Growth Trends for Selected Occupations Considered at Risk from Automation," *Monthly Labor Review*, U.S. Bureau of Labor Statistics, July 2022, https://doi.org/10.21916/mlr.2022.21.

[21]Francesc Pedró et al., *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*, Working Paper on Education Policy 07 (Paris: UNESCO, 2019), https://unesdoc.unesco.org/ark:/48223/pf0000366994.

[22]Lorena Casal-Otero et al., "AI Literacy in K-12: A Systematic Literature Review," *International Journal of STEM Education* 10, no. 1 (2023): 29, https://doi.org/10.1186/s40594-023-00418-7.

[23]For example, Laura Kristen Allen and Panayiota Kendeou, "ED-AI Lit: An Interdisciplinary Framework for AI Literacy in Education," *Policy Insights from the Behavioral and Brain Sciences* 11, no. 1 (2024): 3–10. https://doi.org/10.1177/23727322231220339; "Artificial Intelligence (AI) for K-12 initiative (AI4K12)," ai4k12.org, accessed June 10, 2024, https://ai4k12.org/; "AI 101 for Teachers," Code.org, accessed June 6, 2024, https://code.org/ai/pl/101; Kelly Mills et al., *AI Literacy: A Framework to Understand, Evaluate, and Use Emerging Technology* (Washington DC: Digital Promise, June 2024), https://doi.org/10.51388/20.500.12265/218; Victor Lee and Duri Long, "AI Literacy: Definitions and Directions for an Essential New Digital Literacy," (poster presented at the annual meeting of the American Educational Research Association, Philadelphia, PA, April 2024); Duri Long and Brian Magerko, "What Is AI Literacy? Competencies and Design Considerations," in *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York: ACM, 2020), 1–16, https://doi.org/10.1145/3313831.3376727; Davy Tsz Kit Ng et al., *AI Literacy in K-16 Classrooms* (Cham, Switzerland: Springer, 2023); Davy Tsz Kit Ng et al., "Artificial Intelligence (AI) Literacy Questionnaire with Confirmatory Factor Analysis," in *2023 IEEE International Conference on Advanced Learning Technologies* (ICALT) (New York: IEEE, 2023), 233–235, https://doi.org/10.1109/ICALT58122.2023.00074.

[24]Davy Tsz Kit Ng et al., "Conceptualizing AI Literacy: An Exploratory Review," *Computers and Education: Artificial Intelligence* 2 (2021): 1–27 https://doi.org/10.1016/j.caeai.2021.100041.

[25]Long and Magerko, "What Is AI Literacy?"

[26]Marc Pinski and Benlian Alexander, "AI Literacy—Towards Measuring Human Competency in Artificial Intelligence," in *Proceedings of the 56th Hawaii International Conference on System Sciences,* ed. Tung X. Bui (Honolulu, HI: University of Hawaii at Manoa, 2023), 165–171, https://doi.org/10.24251/HICSS.2023.021.

[27]Omaima Almatrafi, Aditya Johri, and Hyuna Lee, "A Systematic Review of AI Literacy Conceptualization, Constructs, and Implementation and Assessment Efforts (2019–2023)," *Computers and Education Open* 6 (2024): 100173, https://doi.org/10.1016/j.caeo.2024.100173.

[28]Definitions adapted and extended from Almatrafi, Johri, and Lee, "AI Literacy Conceptualization."

[29]Kate Cain, "Literacy Development: The Interdependent Roles of Oral Language and Reading Comprehension," in *Routledge Handbook of Communication Disorders*, eds. Ruth H. Bahr and Elaine R. Silliman (New York, Routledge, 2015), 204–214, https://doi.org/10.4324/9780203569245-22; Maryanne Wolf and Rebecca Kennedy, "How the Origins of Written Language Instruct Us to Teach: A Response to Steven Strauss," *Educational Researcher* 32, no. 2 (2003): 26–30, https://doi.org/10.3102/0013189X032002026.

[30]See Gilster, *Digital Literacy*; Susan R. Goldman and Marlene Scardamalia, "Managing, Understanding, Applying, and Creating Knowledge in the Information Age: Next-Generation Challenges and Opportunities," *Cognition and Instruction* 31, no.2 (2013): 255–269, https://doi.org/10.1080/10824669.2013.773217.

[31]Jason R. Finley, Farah Naaz, and Francine W. Goh, *Memory and Technology: How We Use Information in the Brain and the World* (Cham, Switzerland: Springer, 2018), https://doi.org/10.1007/978-3-319-99169-6.; Roger Säljö, "Literacy, Digital Literacy and Epistemic Practices: The Co-Evolution of Hybrid Minds and External Memory Systems," *Nordic Journal of Digital Literacy* 7, no. 1 (2012): 5–19, https://doi.org/10.18261/ISSN1891-943X-2012-01-02.

[32]Clark A. Chinn, Sarit Barzilai, and Ravit Golan Duncan, "Education for a 'Post-Truth' World: New Directions for Research and Practice," *Educational Researcher* 50, no. 1 (2021): 51–60, https://doi.org/10.3102/0013189X20940683; Rainer Greifeneder et al., "What is New and True about Fake News?" in T*he Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation*, eds. Rainer Greifeneder et al. (Routledge, 2021), 1-8.; Donald J. Leu, Jr. and Charles K. Kinzer, "The Convergence of Literacy Instruction with Networked Technologies for Information and Communication," *Reading Research Quarterly* 35, no. 1 (2000): 108–127, https://doi.org/10.1598/RRQ.35.1.8.

[33]Kenneth R. Koedinger, Albert T. Corbett, and Charles Perfetti, "The Knowledge-Learning-Instruction (KLI) Framework: Toward Bridging the Science-Practice Chasm to Enhance Robust Student Learning," *Cognitive Science* 36, no. 5 (2012): 757–798, https://doi.org/10.1111/j.1551-6709.2012.01245.x.

[34]Joseph P. Magliano et al., "Exploring Thresholds in the Foundational Skills for Reading and Comprehension Outcomes in the Context of Postsecondary Readers," *Journal of Learning Disabilities* 56, no. 1 (2023): 43–57, https://doi.org/10.1177/00222194221087387.

[35]Anastasia Efklides and Bennett L. Schwartz, "Revisiting the Metacognitive and Affective Model of Self-Regulated Learning: Origins, Development, and Future Directions," *Educational Psychology Review* 36, no. 2 (2024): 1–36, https://doi.org/10.1007/s10648-024-09896-9.

[36]See Magliano et al., "Exploring Thresholds"; OECD, *Skills Matter: Additional Results from the Survey of Adult Skills* (Paris: OECD Publishing, 2019), https://doi.org/10.1787/1f029d8f-en.

[37]Chun Chu-Ke and Yujie Dong, "Misinformation and Literacies in the Era of Generative Artificial Intelligence: A Brief Overview and a Call for Future Research," *Emerging Media* 2, no 1 (2024), https://doi.org/10.1177/27523543241240285.

[38]Ziwei Ji et al., "Survey of Hallucination in Natural Language Generation," *ACM Computing Surveys* 55, no. 12 (2023): 1–38, https://doi.org/10.1145/3571730.

[39]Ilaria Amaro et al. "Believe in Artificial Intelligence? A User Study on the ChatGPT's Fake Information Impact," *IEEE Transactions on Computational Social Systems* (2023), https://ieeexplore.ieee.org/abstract/document/10183726; Gairika Mitra, "Glue Pizza, Eat Rocks: Google AI Search Gives Erratic Response," MSN News, June 25, 2024, https://www.msn.com/en-us/news/technology/glue-pizza-eat-rocks-google-ai-search-gives-erratic-responses/ar-BB1n2tcz.

[40]Dragan Gašević, George Siemens, and Shazia Sadiq, "Empowering Learners for the Age of Artificial Intelligence," *Computers and Education: Artificial Intelligence* 4 (2023): https://doi.org/10.1016/j.caeai.2023.100130; Fan Ouyang and Pengcheng Jiao, "Artificial Intelligence in Education: The Three Paradigms," *Computers and Education: Artificial Intelligence* 2 (2021): 100020, https://doi.org/10.1016/j.caeai.2021.100020.

[41]Ido Roll and Ruth Wylie, "Evolution and Revolution in Artificial Intelligence in Education," *International Journal of Artificial Intelligence in Education* 26 (2016): 582–599, https://doi.org/10.1007/s40593-016-0110-3.

[42]OECD, *Technical Report of the Survey of Adult Skills (PIAAC)*, 3rd ed. (Paris: OECD Publishing, 2019), https://www.oecd.org/skills/piaac/publications/PIAAC_Technical_Report_2019.pdf; Sparks, Katz, and O'Neil, *Assessing Digital Information Literacy.*

[43]Sparks and Deane, *Cognitively Based Assessment*; Jesse R. Sparks, "NAEP SAIL: Innovations in Assessing Inquiry Skills Using Virtual Environments" (e-poster presentation at the NAEP Innovations Symposium, Washington, DC, December 2017).

[44]Coiro, Sparks, and Kulikowich, "Assessing Online Collaborative Inquiry"; Sparks, Katz, and O'Neil, *Assessing Digital Information Literacy.*

[45]Vibhor Agarwal and Nishanth Sastry, "'Way Back Then': A Data-Driven View of 25+ Years of Web Evolution," in *WWW '22: Proceedings of the ACM Web Conference*, eds. Frédérique Laforest et al. (New York, ACM, 2022), 3471–3479, https://doi.org/10.1145/3485447.3512283.

[46]Corby Rosset et al., "Researchy Questions: A Dataset of Multi-Perspective, Decompositional Questions for LLM Web Agents," preprint, submitted February 27, 2024, https://doi.org/10.48550/arXiv.2402.17896.

[47]"Where Do DuckDuckGo Search Results Come from?" search results, DuckDuckGo, accessed June 5, 2024, https://duckduckgo.com/duckduckgo-help-pages/results/sources/.

[48]Yutao Zhu et al., "Large Language Models for Information Retrieval: A Survey," preprint, last modified January 19, 2024, https://doi.org/10.48550/arXiv.2308.07107.

[49]Rosset et al., "Researchy Questions."

[50]Sofia Eleni Spatharioti et al., "Comparing Traditional and LLM-Based Search for Consumer Choice: A Randomized Experiment," preprint, last modified November 8, 2023, https://doi.org/10.48550/arXiv.2307.03744.

[51]Sunnie S.Y. Kim et al. "'I'm Not Sure, But... ': Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust," in *2024 ACM Conference on Fairness, Accountability, and Transparency* (New York: ACM, 2024), https://doi.org/10.1145/3630106.3658941.

[52]Nikhil Sharma, Q. Vera Liao, and Ziang Xiao, "Generative Echo Chamber? Effect of LLM-Powered Search Systems on Diverse Information Seeking," in *CHI '24: Proceedings of the CHI Conference on Human Factors in Computing Systems*, eds. Florian Floyd Mueller et al. (New York: ACM, 2024), 1-17, https://doi.org/10.1145/3613904.3642459.

[53]See "OpenAI: GPT-4," OpenAI, published March 14, 2023, https://openai.com/index/gpt-4-research/

[54]See Daniel L. Schwartz and Dylan Arena, *Measuring What Matters Most: Choice-Based Assessments for the Digital Age* (Cambridge: The MIT Press, 2013), 192, https://doi.org/10.7551/mitpress/9430.001.0001.

[55]Spatharioti et al., "Comparing Traditional and LLM-Based Search."

[56]Joshua Maynez et al., "On Faithfulness and Factuality in Abstractive Summarization," preprint, submitted May 2, 2020, https://doi.org/10.48550/arXiv.2005.00661; Ji et al., "Survey of Hallucination."

[57]Janek Bevendorff et al., "Is Google Getting Worse? A Longitudinal Investigation of SEO Spam in Search Engines," in *Advances in Information Retrieval: 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24–28, 2024, Proceedings*, Part III, eds. Nazli Goharian et al. (Cham, Switzerland: Springer, 2024), 56–71, https://doi.org/10.1007/978-3-031-56063-7_4.

[58]See Sarit Barzilai, Eynav Tzadok, and Yoram Eshet-Alkalai, "Sourcing while reading divergent expert accounts: Pathways from views of knowing to written argumentation," *Instructional Science* 43 (2015), 737–766, https://doi.org/10.1007/s11251-015-9359-4; Ivar Bråten, Helge I. Strømsø, and Ladislao Salmeron, "Trust and mistrust when students read multiple information sources about climate change," *Learning and Instruction* 21, no. 2 (2013), 180–192, https://doi.org/10.1016/j.learninstruc.2010.02.002; Helge I. Strømsø et al., "Spontaneous Sourcing Among Students Reading Multiple Documents," *Cognition and Instruction* 31, no. 2 (2013): 176–203, https://doi.org/10.1080/07370008.2013.769994.

[59]See Coiro, Sparks, and Kulikowich, "Assessing Online Collaborative Inquiry"; Sparks and Deane, *Cognitively Based Assessment*; Tenison and Sparks, "Combining Cognitive Theory."

[60]Jennifer Wiley and James F. Voss, "Constructing Arguments from Multiple Sources: Tasks that Promote Understanding and not Just Memory for Text," *Journal of Educational Psychology* 91, no. 2 (1999): 301–311, https://doi.org/10.1037/0022-0663.91.2.301.

[61]M. Anne Britt and Jean-François Rouet, "Learning with Multiple Documents: Component Skills and their Acquisition," in *Enhancing the Quality of Learning: Dispositions, Instruction, and Learning Processes*, eds. John R. Kirby and Michael J. Lawson (New York: Cambridge University Press, 2012), 276–314, https://doi.org/10.1017/CBO9781139048224.017; Britt, Rouet, and Durik, *Literacy beyond Text Comprehension*.

[62]Maynez et al., "On Faithfulness."

[63]Spatharioti et al., "Comparing Traditional and LLM-Based Search."

[64]Sharma, Liao, and Xiao, "Generative Echo Chamber?"

[65]Maurice Jakesch et al., "Co-writing with opinionated language models affects users' views," in *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems,* eds. Albrecht Schmidt et al. (New York: ACM, 2023), 1-15, https://doi.org/10.1145/3544548.3581196.

[66]We note that this kind of susceptibility to belief in information one has read or engaged with is not a phenomenon unique to LLMs (see the following papers by Daniel Gilbert: Daniel T. Gilbert, "How Mental Systems Believe," *American Psychologist* 46, no. 2 (1991): 107–119, https://doi.org/10.1037/0003-066X.46.2.107; Daniel T. Gilbert, Douglas S. Krull, and Patrick S. Malone, "Unbelieving the Unbelievable: Some Problems in the Rejection of False Information," *Journal of Personality and Social Psychology* 59, no. 4 (1990): 601–613, https://doi.org/10.1037/0022-3514.59.4.601; Daniel T. Gilbert, Romin, W. Tafarodi, and Patrick S. Malone, "You Can't Not Believe Everything You Read," *Journal of Personality and Social Psychology* 65, no. 2 (1993): 221–233, https://doi.org/10.1037/0022-3514.65.2.221), but may be exacerbated by their use.

[67]Margaret Bearman, Juuso Henrik Nieminen, and Rola Ajjawi, "Designing Assessment in a Digital World: An Organising Framework," *Assessment & Evaluation in Higher Education* 48, no. 3 (2023): 291–304, https://doi.org/10.1080/02602938.2022.2069674.

[68]Robert J. Mislevy, Russell G. Almond, and Janice F. Lukas, *A Brief Introduction to Evidence-Centered Design*, Research Report No. RR-03-16 (Princeton, NJ: ETS, 2003), https://doi.org/10.1002/j.2333-8504.2003.tb01908.x; Robert J. Mislevy, Linda S. Steinberg, and Russell G. Almond, "Focus Article: On the Structure of Educational Assessments," *Measurement: Interdisciplinary Research and Perspectives* 1, no. 1 (2003): 3–62, https://doi.org/10.1207/S15366359MEA0101_02.

[69]American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for Educational and Psychological Testing* (Washington, DC: American Educational Research Association, 2014).

[70]Mislevy, Almond, and Lukas, *Evidence-Centered Design*; Mislevy, Steinberg, and Almond, "Structure of Educational Assessments."

[71]Robert J. Mislevy, *Sociocognitive Foundations of Educational Measurement* (New York: Routledge, 2018), https://doi.org/10.4324/9781315871691.

[72]Mislevy, Almond, and Lukas, *Evidence-Centered Design*; Mislevy, Steinberg, and Almond, "Structure of Educational Assessments."

[73] See Deirdre Kerr, Jessica J. Andrews, and Robert J. Mislevy, "The In-Task Assessment Framework for Behavioral Data," in *The Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications*, eds. A. A. Rupp and J. P. Leighton (Hoboken, NJ: Wiley-Blackwell, 2016): 472–507. https://doi.org/10.1002/9781118956588.ch20.

[74]See Luci Pangrazio and Julian Sefton-Green, "Digital Rights, Digital Citizenship and Digital Literacy: What's the Difference?" *Journal of New Approaches in Educational Research* 10, no. 1 (2021): 15–27, https://doi.org/10.7821/naer.2021.1.616; Sparks, Katz, and O'Neil, *Assessing Digital Information Literacy*.

[75] Michael T. Kane, "Validating the Interpretations and Uses of Test Scores," *Journal of Educational Measurement* 50, no. 1 (2013): 1–73, https://doi.org/10.1111/jedm.12000.

[76]See Marcus O'Donnell, "Assessment as and of Digital Practice: Building Productive Digital Literacies," in *Re-Imagining University Assessment in a Digital World*, eds. Margaret Bearman et al. (Cham, Switzerland: Springer, 2020), 111–125, https://doi.org/10.1007/978-3-030-41956-1_9.

[77]See Pangrazio, Godhe, and Ledesma, "What Is Digital Literacy?"

[78] OECD, *PISA 2021 ICT Framework* (Paris: OECD Publishing, 2019), https://www.oecd.org/pisa/sitedocument/PISA-2021-ICT-framework.pdf.

[79]Julian Fraillon et al., *IEA International Computer and Information Literacy Study 2018 Assessment Framework* (Cham, Switzerland: Springer, 2018), 13–23, https://doi.org/10.1007/978-3-030-19389-8_2; Nancy Law et al., *A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2* (Montreal, Quebec: UNESCO Institute for Statistics, 2018), https://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf.

[80]See UNESCO, *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*, Working Papers on Education Policy 07 (Paris: UNESCO, 2019), https://unesdoc.unesco.org/ark:/48223/pf0000366994/PDF/366994eng.pdf.multi.

[81]See Sparks and Deane, Cognitively Based Assessment; Jesse R. Sparks, Peter W. van Rijn, and Paul Deane, "Assessing Source Evaluation Skills of Middle School Students Using Learning Progressions," *Educational Assessment* 26, no. 4 (2021): 213–240, https://doi.org/10.1080/10627197.2021.1966299.

[82]Ng et al., "Conceptualizing AI Literacy"; Martin Kandlhofer et al., "Artificial Intelligence and Computer Science in Education: From Kindergarten to University," in *2016 IEEE Frontiers in Education Conference (FIE)* (Red Hook, NY Curran & Associates, Inc., 2016), 1–9, https://doi.org/10.1109/FIE.2016.7757570.

[83]Weipeng Yang, "Artificial Intelligence Education for Young Children: Why, What, and How in Curriculum Design and Implementation," *Computers and Education: Artificial Intelligence* 3 (2022): 100061, https://doi.org/10.1016/j.caeai.2022.100061.

[84]See Table 4.8 in Davy Tsz Kit Ng et al., "AI Literacy for All," in *AI Literacy in K-16 Classrooms*, eds. Davy Tsz Kit Ng, Jac Ka Lok Leung, Maggie Jiahong Su, Iris Heung Yue Kim, Maggie Shen Qiao, and Samuel Kai Wah Chu (Cham: Springer International Publishing, 2022), 21-29, https://doi.org/10.1007/978-3-031-18880-0_3.

[85]See Coiro, Sparks, and Kulikowich, "Assessing Online Collaborative Inquiry"; Tenison and Sparks, "Combining Cognitive Theory."

[86]John Sabatini et al., "Engineering a 21st Century Reading Comprehension Assessment System Utilizing Scenario-Based Assessment Techniques," *International Journal of Testing* 20, no. 1 (2019): 1–23, https://doi.org/10.1080/15305058.2018.1551224.

[87]Rola Ajjawi et al. "From Authentic Assessment to Authenticity in Assessment: Broadening Perspectives," *Assessment & Evaluation in Higher Education* 49, no. 4 (2023): 499–510, https://doi.org/10.1080/02602938.2023.2271193; Sparks and Deane, *Cognitively Based Assessment*.

[88]Coiro, Sparks, and Kulikowich, "Assessing Online Collaborative Inquiry"; Paul Deane and Jesse R. Sparks, "Scenario-Based Formative Assessment of Key Practices in the English Language Arts," in *Handbook of Formative Assessment in the Disciplines*, eds. Heidi L. Andrade, Randy E. Bennett, and Gregory J. Cizek (New York: Routledge, 2019), 68–96, https://doi.org/10.4324/9781315166933; Sparks and Deane, *Cognitively Based Assessment*; Tenison and Sparks, "Combining Cognitive Theory."

[89]Schwartz and Arena, *Measuring What Matters Most*.

[90]Ido Roll and Miri Barhak-Rabinowitz, "Measuring Self-Regulated Learning Using Feedback and Resources," in *Innovating Assessments to Measure and Support Complex Skills*, eds. Natalie Foster and Mario Piacentini (Paris: OECD Publishing, 2023), https://doi.org/10.1787/c93ac64e-en.

[91]Diego Zapata-Rivera et al., "Designing and Evaluating Evidence-Centered Design Based Conversations for Assessment with LLMs," in *Proceedings of EDM 2024 Workshop: Leveraging Large Language Models for Next Generation Educational Technologies*, forthcoming.

[92]Madeleine Keehner, Burcu Arslan, and Marlit Annalena Lindne, "Cognition-Centered Design Principles for Digital Assessment Tasks and Items," in *International Encyclopedia of Education*, 4th ed., eds. Robert J Tierney, Fazal Rizvi, and Kadriye Ercikan (Amsterdam: Elsevier, 2023), 171–184, https://doi.org/10.1016/B978-0-12-818630-5.10025-9.

[93]Keehner, Arslan, and Lindner, "Cognition-Centered Design Principles."

[94]See Burcu Arslan, Caitlin Tenison, and Bridgid Finn, "Going beyond Observable Actions: A Cognition-Centered Approach to Interpreting Pauses Represented in Process Data," *European Journal of Psychological Assessment* 39, no. 4 (2023): 263–270, https://doi.org/10.1027/1015-5759/a000756.

[95]See, for example, Tenison and Sparks, "Combining Cognitive Theory."

[96]Mislevy, Steinberg, and Almond, "Structure of Educational Assessments."

[97]Tenison and Sparks, "Combining Cognitive Theory."

[98]See Christine DeMars, *Item Response Theory* (New York: Oxford University Press, 2010), https://doi.org/10.1093/acprof:oso/9780195377033.001.0001.

[99]Sparks, Katz, and O'Neil, *Assessing Digital Information Literacy*.

[100]Roll and Barhak-Rabinowitz, "Measuring Self-Regulated Learning."

[101]Diego Zapata-Rivera, ed., *Score Reporting Research and Applications* (New York: Routledge, 2019).

[102]See Steven L. Wise and Lisa F. Smith, "A Model of Examinee Test-Taking Effort," in *High-Stakes Testing in Education: Science and Practice in K-12 Settings*, eds. James A. Bovaird, Kurt F. Geisinger, and Chad W. Buckendahl (Washington, DC: American Psychological Association, 2011), 139–153, https://doi.org/10.1037/12330-009; Steven L. Wise and Lisa F. Smith, "The Validity of Assessment When Students Don't Give Good Effort," in *Handbook of Human and Social Conditions in Assessment*, eds. Gavin T. L. Brown and Lois R. Harris (New York: Routledge, 2016), 204–220.

[103]Diego Zapata-Rivera and Irvin R. Katz, "Keeping Your Audience in Mind: Applying Audience Analysis to the Design of Interactive Score Reports," *Assessment in Education: Principles, Policy & Practice* 21, no. 3 (2014): 442–463, https://doi.org/10.1080/0969594X.2014.936357.

[104]Priya Kannan and Diego Zapata-Rivera, “Facilitating the Use of Data from Multiple Sources for Formative Learning in the Context of Digital Assessments: Informing the Design and Development of Learning Analytic Dashboards,” *Frontiers in Education* 7 (2022), https://doi.org/10.3389/feduc.2022.913594; Carol M. Forsyth et al., “Interactive Score Reporting: An AutoTutor-Based System for Teachers,” in *Artificial Intelligence in Education: 18th International Conference, AIED 2017, Wuhan, China, June 28–July 1, 2017, Proceedings*, ed. Elisabeth André et al. (Cham, Switzerland: Springer, 2017), 506–509, https://doi.org/10.1007/978-3-319-61425-0_51; Diego Zapata-Rivera, “Adaptive Score Reports,” in *User Modeling, Adaptation, and Personalization: 20th International Conference, UMAP 2012, Montreal, Canada, July 16-20, Proceedings*, eds. Judith Masthoff et al. (Cham, Switzerland: Springer, 2012), 340–345, https://doi.org/10.1007/978-3-642-31454-4_32; Diego Zapata-Rivera and Burcu Arslan, “Learner Modeling Interpretability and Explainability in Intelligent Adaptive Systems,” in Mind, Body, and *Digital Brains*, eds. Flavia Santoianni, Gianluca. Giannini, and Alessandro Ciasullo (Cham, Switzerland: Springer, forthcoming).

[105]Ng et al., “Conceptualizing AI Literacy,” 41.

[106]For example, Fraillon et al., *IEA Computer and Information Literacy Study*; OECD, *PISA 2021 ICT Framework*; Nancy Law et al., *Framework on Digital Literacy Skills*.

[107]UNESCO, *K-12 AI Curricula: A Mapping of AI Curricula* (Paris: UNESCO, 2022), https://unesdoc.unesco.org/ark:/48223/pf0000380602.

[108]See OECD, *PISA 2018 Assessment and Analytical Framework* (Paris: OECD Publishing, 2019), https://www.oecd-ilibrary.org/education/pisa-2018-assessment-and-analytical-framework_b25efab8-en; OECD, *Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots* (Paris: OECD Publishing, 2021), https://www.oecd-ilibrary.org/education/oecd-digital-education-outlook-2021_589b283f-en.

[109]Alexander J. van Deursen and Jan A. G. M. van Dijk, “Toward a Multifaceted Model of Internet Access for Understanding Digital Divides: An Empirical Investigation,” *The Information Society* 31, no. 5 (2015): 379–391, https://doi.org/10.1080/01972243.2015.1069770.

[110]See Pangrazio, Godhe, and Ledesma, “What Is Digital Literacy?” for comparisons of digital literacies constructs across English, Spanish, and Scandinavian language contexts.

[111]Pangrazio, Godhe, and Ledesma “What Is Digital Literacy?”

[112]See Ng et al., “Conceptualizing AI Literacy.”

[113]See Future of Life Institute, *The EU Artificial Intelligence* Act, (2024), https://artificialintelligenceact.eu/ai-act-explorer/.

[114]Francesca Borgonovi et al., *Emerging Trends in AI Skill Demand across 14 OECD Countries*, OECD Artificial Intelligence Papers No. 2 (Paris: OECD Publishing, 2023), https://doi.org/10.1787/7c691b9a-en.

[115]See Samuel Messick, “The Interplay of Evidence and Consequences in the Validation of Performance Assessments,” *Educational Researcher* 23, no. 2 (1994): 13–23, https://doi.org/10.3102/0013189X023002013.

[116] Steven Lewis and Bob Lingard, “The Multiple Effects of International Large-Scale Assessment on Education Policy and Research,” *Discourse: Studies in the Cultural Politics of Education* 36, no. 5 (2015): 621–637, https://doi.org/10.1080/01596306.2015.1039765.

[117] See Roll and Barhak-Rabinowitz, “Measuring Self-Regulated Learning.”

[118]Rebecca. J. Passonneau et al., “Preface: Special Issue on Multidisciplinary Approaches to AI and Education for Reading and Writing,” *International Journal of Artificial Intelligence in Education* 27 (2017): 665–670, https://doi.org/10.1007/s40593-017-0158-8.

[119]See van Deursen and van Dijk, “Multifaceted Model of Internet Access.”

[120]See Bearman, Nieminen, and Ajjawi, “Designing Assessment.”

[121]See Corrine M. Wickens and Tracy Miller, “Gender, Digital Literacies, and Higher Education: Examinations of Equity,” in *The Wiley Handbook of Gender Equity in Higher Education*, eds. Nancy S. Niemi and Marcus B. Weaver-Hightower (New York: Routledge, 2020), 53–67, https://doi.org/10.1002/9781119257639.ch3.

[122]Magliano et al., “Exploring Thresholds”; Tenaha O’Reilly, Zuowei Wang, and John Sabatini, “How Much Knowledge Is Too Little? When Knowledge Becomes a Barrier to Comprehension,” *Psychological Science* 30, no. 29 (2019): 1344–1351, https://journals.sagepub.com/doi/10.1177/0956797619862276; Zuowei Wang et al., “Decoding and Reading Comprehension: A Test of the Decoding Threshold Hypothesis,” *Journal of Educational Psychology* 111, no. 3 (2019): 387–401, https://psycnet.apa.org/doiLanding?doi=10.1037%2Fedu0000302.

[123]J. D. Zamfirescu-Pereira et al., “Why Johnny Can’t Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts,” in *CHI ‘23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, eds. Albrecht Schmidt et al. (New York: ACM, 2023), 1–21, https://doi.org/10.1145/3544548.3581388.

[124]Nils Knoth et al., “AI Literacy and Its Implications for Prompt Engineering Strategies,” *Computers and Education: Artificial Intelligence* 6 (2024): 100225, https://doi.org/10.1016/j.caeai.2024.100225.

[125]See Gonzalo Martínez et al., “Towards Understanding the Interplay of Generative Artificial Intelligence and the Internet,” in E*pistemic Uncertainty in Artificial Intelligence: First International Workshop, Epi UAI 2023, Pittsburgh, PA, USA, August 4, 2023, Revised Selected Papers*, eds. Fabio Cuzzolin and Maryam Sultana (Cham, Switzerland: Springer, 2023), 59–73. https://doi.org/10.1007/978-3-031-57963-9_5.

[126]See Guillermo Solano-Flores, “Examining Cultural Responsiveness in Large-Scale Assessment: The Matrix of Evidence for Validity Argumentation,” *Frontiers in Education* 4 (2019), https://doi.org/10.3389/feduc.2019.00043.

[127]David Pagliaccio et al., “Probing the Digital Exposome: Associations of Social Media Use Patterns with Youth Mental Health,” *NPP—Digital Psychiatry and Neuroscience* 2, no. 1 (2024): 5, https://doi.org/10.1038/s44277-024-00006-9.