Assessment of adults with low literacy skills: a review of methods

Katarzyna Chyl[1]*, Agnieszka Dębska[2], Artur Pokropek[1], Marcin Szczerbiński[3], Łukasz Lech Tanaś[4] and Michał Sitek[1]

[1]International Projects Unit, Educational Research Institute, Warsaw, Poland, [2]Laboratory of Language Neurobiology, Nencki Institute of Experimental Biology (PAS), Warsaw, Poland, [3]School of Applied Psychology, University College Cork, Cork, Ireland, [4]Faculty of Psychology, University of Social Sciences and Humanities, Warsaw, Poland

The paper reviews the methods for assessing different components of reading skills in adults with reading difficulties, along with *functional* reading skills. We are particularly interested in the assessment methods available to researchers and practitioners, developed predominantly in the research context, and not available solely in English. We discuss the large-scale international study, PIAAC, as an example of a framework for such assessments. Furthermore, we cover the following types of assessment tools: (1) self-assessment questionnaires, probing into comprehension difficulties and reading habits; (2) measures of print exposure, such as author recognition tests, correlating with other reading-related skills; (3) measures of word recognition and decoding, including reading aloud of words and pseudowords, as well as silent lexical decision tasks; (4) fill-in-the-blank tasks and sentence reading tasks, measuring predominantly local comprehension, entangled with decoding skills; (5) comprehension of longer reading passages and texts, focusing on functional texts. We discuss comprehension types measured by tests, text types, answer formats, and the dependence problem, i.e., reading comprehension tests that can be solved correctly without reading. Finally, we tap into the new ideas emerging from the AI systems evaluation, e.g., using questions generated from news articles or Wikipedia or asked directly by search engines users. In the concluding section, we comment on the significance of incorporating background information, motivation, and self-efficacy into the assessment of adult literacy skills.

Introduction

In this review, we focus on evaluating the lower spectrum of adult reading skills. When delving into this topic, it's crucial to consider two fundamental views of literacy:

1. Literacy as a set of cognitive skills related to recovering language information from print. This encompasses two aspects of cognitive processing, namely converting print to speech sounds (decoding, word recognition, phonological awareness, automatization) and recovering meaning (lexical access, sentence parsing, building mental models, etc. (e.g., Hoover and Gough, 1990).
2. Functional literacy, which refers to the proficiency needed to navigate the demands of a literate society, i.e., applying reading in context. It involves tasks like understanding instructions, interpreting documents, filling out forms, and making informed decisions based on written information (Vágvölgyi et al., 2016).

These two narratives of thinking about literacy emerged within the context of different educational practices and different scientific disciplines; they also frame our key problem of measuring reading skills differently. "Literacy as a cognitive skill" narrative is associated with experimental psychology and traditional educational practices. Measurement-wise, this narrative is epitomized in the vast array of psychometric tools, where reading is broken down into its constituent components, each measured separately. These components are cognitively rather than functionally salient (e.g., sight word reading). The measurement, if done well, is meant to give us a comprehensive description of the cognitive architecture of the reading process, as it is executed in the reader's mind. It also covers reading-related skills, such as vocabulary, or phonological skills. Tests usually provide norms, based on the population scores.

The term "functional literacy" is often left undefined in the literature (Perry et al., 2018). If the definitions are used, they are typically institutionally driven (e.g., OECD, 2012). Historically, this term was coined during World War II to indicate soldiers who were unable to use written instructions to adequately perform basic military tasks (Castell et al., 1981). In the second half of the 20th century, functional literacy gained interest from policymakers and researchers (Vágvölgyi et al., 2016). The working definitions underlined the "real-life" aspect of literacy skill use, and they were "survival-oriented". They tapped into job-seeking, transportation, and economic necessities. Skills considered "functional" were defined as those used to obtain food, clothing, healthcare, etc. (Kirsch and Guthrie, 1977). In the past, arbitrary thresholds were proposed for literacy to be considered "functional", such as years of schooling. However, this criterion is arbitrary, as the number of years of education deemed necessary to become functionally literate varied across decades and countries. There is also abundant evidence that a significant proportion of people who completed many years of education have low functional literacy. Another diagnostic practice is to use grade-equivalent scores, where low literate adults are compared with school-age children. While useful, this approach ignores differences in adults and children's contexts and developmental stages (Vágvölgyi et al., 2016).

Currently, UNESCO and OECD have stopped offering a unitary definition of "functionality" of literacy (Vágvölgyi et al., 2016). But following the 1978 UNESCO General Conference, researchers still use the negative definition: inability to understand, evaluate, use, and engage in the written text to participate in society, achieve goals, and develop knowledge and potential. In practice, the creators of the prominent Program for the International Assessment of Adult Competencies (PIAAC) decided that achieving the third (out of six) level of text comprehension is required to reach the point of "participating in society". Reading comprehension tasks below this level are limited to short texts, locating single pieces of information, and making the simplest of inferences. Even though in the high-income countries illiteracy is no longer a social problem, still around 15% of adults assessed in the first cycle of PIAAC are placed below the "functional" level (OECD, 2012).

Most comparative literacy surveys are children- and school-related, as the recent bibliometric review has shown (Lan and Yu, 2022). Studies that are predominantly interested in adults are scarce. In the international large-scale assessments, many more countries participated in school-age assessments (PISA 2022: 81, PIRLS 2021: 57) than those focused on adults (PIAAC 2008–2019: 33, PIAAC 2023: 31). It is also interesting that the adult-focused assessments tend to be organized specifically in low- and middle-income countries (World's Bank Skills Measurement Program STEP 2012–2017: 17, UNESCO's Literacy Assessment and Monitoring Program: 5), where average literacy levels may be lower. Some studies also operate at the country level and test their citizens' literacy skills (e.g., the German National Education Panel Study NEPS), which might be especially useful for policymakers.

Applied contexts of literacy measurement

Adult literacy can be assessed through gathering standardized data from large, nationally representative samples, and doing so repeatedly. This can then be used for cross-national comparisons or within-country longitudinal comparisons. Such assessment gained broader recognition with a series of studies organized since the mid-1990s by the OECD: the International Adult Literacy Survey (IALS 1994–1998), the Adult Literacy and Life Skills Survey (ALL 2002–2008), and the Program for the International Assessment of Adult Competencies, 2008–2019 and ongoing second cycle. Large-scale assessments describe populations of interest; in the case of the 2nd cycle of PIAAC, in the areas of literacy, numeracy, and problem-solving. Consequently, these measurements focus on country means and performance distributions. Importantly, the test materials are designed to be used solely within the context of this research panel and are not made available to the practitioners (educational psychologists, etc.). The ultimate goal is to compare groups, e.g., from different countries, to understand how the skills of interest relate to educational, economic, and social outcomes (Kirsch and Lennon, 2017).

Beyond comparative surveys, adult literacy skills are commonly assessed in the context of job recruitment, adult formal education, or social-institutional support (Murray, 2017). Undertaking the assessment could help a person determine if the reading level suits their goals—e.g., meets job requirements or college class level. Current reading level diagnosis might suggest the appropriate type and level of training necessary to meet the goals. Later on, in the course of the training, the assessment could determine the progress made and identifies gaps that still need to be filled. In some higher-income countries, an adult with lower literacy skills might look for a course, tutoring, or training in a specialized service, e.g., the Adult Basic Education (ABE) Program in the United States; Education and Training Boards (ETB) in Ireland; National Literacy Trust and Adult Literacy Trust in the UK. Sometimes, enrollment in the course might involve an official reading assessment. For example, in the US it might be the Test for Adult Basic Education (TABE), Comprehensive Adult Student Assessment System (CASAS), Wonderlic General Assessment of Instructional Needs (GAIN), or the Massachusetts Adult Proficiency Tests (MAPT).

What is common for both types of uses mentioned above is the relative public unavailability of test materials. This has an economical (copyright) and procedural justification but also can hinder the development of basic research or innovative assessment

methods in this area. The other reason is the scarcity of suitable tools. A common practice, among both researchers and literacy educators, is the usage of children's tests. However, even if there are some similarities between the reading profile of young children and adults with low literacy skills (Barnes and Kim, 2016), the differences are far more important. Adults use different strategies than children, e.g., relying less on phonology and more on remembering words patterns (Thompkins and Binder, 2003), or more on prior knowledge when reading texts (Greenberg et al., 2009). The differences are also found at the brain level (Martin et al., 2015). Different life experiences, personal interests, and group homogeneity make the comparison between young children and adults with low reading skills very difficult. If we are careful enough to develop separate tests and norms for different school grades, we should do the same for adults, at least for these reading-related skills that go beyond decoding and sentence reading.

The current paper offers a review of available tools for literacy assessment in adults. Some assessments we discuss have norms reaching only young adulthood but are used for adult reading research nonetheless—we discuss those too. It is certainly not an exhaustive list but offers an overview of the types of instruments used in this field. The tools listed here are generally available for practitioners or at least well-known from the English literature, and they were predominantly developed in the context of basic research.

Self-assessment questionnaires

The intuitive view of literacy assessment might work on the assumption that we can "just ask" whether someone experiences difficulties in reading. This approach can be considered a functional assessment—an individual is asked how they cope in everyday life with their reading skills. Examples of such self-assessment reading questionnaires are presented in Table 1. We focused on the published tools available for the larger audience and described in the research papers, selecting an arbitrary cut-off at the year 2000.

This approach receives less attention than formal skill testing, for two reasons (Perry et al., 2017). First, reading is not easily measured using metacognition (Sticht, 2001), with subjects rarely having access to a fine-grained understanding of their literacy skills in different contexts (Boudard and Jones, 2003). It is therefore possible that people simply do not know how much of the "publicly available literacy information" they do not understand (OECD/Statistics Canada, 1995; for a critical analysis, see Hamilton and Barton, 2000). Second, questions about "literacy" are heavily loaded with social desirability, i.e., the tendency to answer questions in a manner that others will view favorably (Olson et al., 2011). As a consequence, results from the parallel self-assessments are more optimistic than performance-based scores (Hautecoeur, 2000; Sticht, 2001). Additionally, as Murray (2017) notes, the relationship between self-perceived literacy and performance-based assessment varies among subpopulations within countries, across countries, and over time, which sums up the limited usability of self-assessments in practice. There are subgroups within populations that are more accurate than others in self-assessments, e.g., middle-aged and females (Gilger, 1992).

Recently, there has been a trend toward shorter self-report questionnaires, as indicated by the publication of versions with fewer items. Morris et al. (2006) created a Single Item Literacy Scale. He analyzed 36 items from the Short Test of Functional Health Literacy in Adults (S-TOFHLA) to select just one question, which best predicts the summary score of the whole scale: "*How often do you need to have someone help you when you read instructions, pamphlets, or other written material from your doctor or pharmacy*?". Brice et al. (2014) created a Two Item Literacy Scale, also based on S-TOFHLA, which used two questions. One asked about the last grade completed and the other about self-estimated level of reading skills, ranging from "reading complete books" to "needing help with newspapers". Both questionnaires had similar characteristics in predicting S-TOFHLA (Baker et al., 1999) scores and detecting readers with lower literacy skills. The benefit of this approach lies in the fact that for general screening purposes, we usually require satisfactory and easy-to-use tools.

The language complexity of the questionnaires themselves is rarely measured, for example, out of the self-report questionnaires below, none reported any measure of word difficulty or sentence length. Questionnaires of high language complexity, which are aimed at literacy assessment, are by definition biased, favoring better readers (Atcherson et al., 2013; Patalay et al., 2018). Many frameworks for assessing language complexity are available, such as Flesch-Kincaid Reading Grade Level (approximate reading grade level of a text), or Gunning Fog Index (estimating the years of formal education a person needs to understand the text). It is advisable to control for this aspect when developing self-report questionnaires.

On the positive side, self-report questionnaires should be most useful for assessment of subjective variables, such as reading motivation, self-efficacy or perception of stress related to non-functional literacy skill. And since reading motivation may be the most important factor for adult readers, this method of assessment requires attention (Frijters et al., 2019). Development of new measures, which would assess the level of literacy-related limitations which an individual encounters in daily life, is also warranted. The current approach of defining the cutoff point for "functional" literacy is heavily dependent on the opinions of test-creators, and not on subjective evaluation. Individuals should be more involved, as "experts in their cause", in creation of such measures and in the process of literacy diagnosis.

Performance-based tests
Measures of print exposure

The bidirectional relationship between reading skills and reading habits, as well as concerns about social desirability bias, justify the construction of indirect measures of print exposure which adopt recognition format. In such tests, respondents are asked to read through the list of authors' names, book titles, titles of periodicals or daily newspapers, and indicate which are real and which are fake (where half of all items are foils). The first of such checklists, the Author Recognition Test and Magazine Recognition Test (Stanovich and West, 1989) were

TABLE 1 Self-assessment questionnaires of reading and dyslexia for adults.

Name of the questionnaire	References	Language	Component measured
Adult reading motivation scale	Schutte and Malouff, 2007	German	Reading motivation
Habitual Reading Motivation Questionnaire	Möller and Bonerad, 2007	German	Reading motivation
Reading habits and attitudes	Applegate et al., 2014	English	Reading habits
Self-Report Habit Index for Reading (SRHI-R)	Schmidt and Retelsdorf, 2016	German	Reading habits
Self-report questionnaire on reading-writing difficulties (ATLAS)	Giménez et al., 2015	Spanish	Reading and writing
Two Item Literacy Screener (TILS)	Brice et al., 2014	English	Reading
Diagnostic Literacy Scale (DIS)	De Greef et al., 2013	English	Reading
Self-testing questionnaire for reading difficulties (ADDA)	Corredor et al., 2019	Spanish	Reading
Single Item Literacy Screener (SILS)	Morris et al., 2006	English	Reading
Self Assessed Literacy Index (SALI)	Olson et al., 2011	English	Reading
Reading Difficulties Questionnaire	Yagyu et al., 2021	Japanese	Reading
Adult Reading Questionnaire (ARQ)	Snowling et al., 2012	English	Dyslexia
Adult Reading History Questionnaire—Brief	Feng et al., 2022	English	Dyslexia
Adult Reading History Questionnaire	Lefly and Pennington, 2000	English	Dyslexia
Adult dyslexia checklist	Stark et al., 2023	English	Dyslexia
Group screening of dyslexia	Wolff and Lundberg, 2003	Swedish	Dyslexia

designed to avoid the social desirability bias burdening the self-assessment questionnaires; especially college students, examined in their study, believed that "it is a good thing to read". Together with the checklists, the authors constructed the Reading and Media Habits questionnaire. The Adult Recognition Test turned out to correlate significantly with decoding skills, word naming, reading comprehension, orthographic processing, and phonological processing (Stanovich and West, 1989), while the Reading and Media Habits' questionnaire did not. Measures of author recognition were also predictive of vocabulary knowledge—much better than reading self-report (Krashen and Kim, 1998) and general language competence. The latter relationship appears specific to the recognition of the authors of fiction (literature), rather than non-fiction (expository) work (Rain and Mar, 2014; Mar and Rain, 2015). Control measures of familiarity with TV programs, TV personalities, movie titles, actors showed weaker or negligible correlation with literacy skills (West et al., 1993).

Why are author recognition measures predictive of language and literacy outcomes?. The knowledge of whether E. L. James is the real name of an author seems to be an arbitrary bit of culturally specific knowledge (Moore and Gordon, 2015). Stanovich and West (1989) suggest that this knowledge was most likely acquired through reading. It is also a manifestation of the literacy environment of a person and their cultural capital (Bourdieu, 1991), similar to, e.g., a question about the number of books at home or parental education. Certainly, extracurricular book reading causes better decoding, a vice versa. Adults who spent time reading while waiting in the departure lounge of an airport were better at decoding than those who occupied themselves otherwise (West et al., 1993); reading for pleasure affects reading skills, which

in turn affects the degree to which reading is a pleasant experience (Moore and Gordon, 2015).

Since the creation of the Author Recognition Test and Magazine Recognition Test, several language versions have been created (e.g., Chinese: Chen and Fang, 2015; Korean: Lee et al., 2019; Dutch: Brysbaert et al., 2020). This is welcomed; even proficient English L2 speakers can't be reliably assessed with the English version of the task (McCarron and Kuperman, 2021). The print exposure measurement paradigm remains prolific in English too—English versions of both Author and Magazine Recognition Tests were updated according to more recent bestsellers (Acheson et al., 2008), and re-designed to directly compare exposure to fiction and non-fiction authors (Mar and Rain, 2015). Indeed, such regular updates will be necessary to keep the measures valid.

As a sidenote, another indirect measure of print exposure is the ability to read low frequency irregular words (such as *aisle*, *chord*, *dept*, or *colonel*) correctly (e.g., National Adult Reading Test, NART; Nelson and Willison, 1991). The assumption behind such tests is that, since such words cannot be decoded correctly using standard grapheme-to-phoneme conversion rules, they must be recognized, i.e., be known to the reader. This implies high print exposure and also high crystallized verbal intelligence. In languages with more transparent orthography, different approaches needed to be taken; for example, the Polish Adult Reading Test (PART; Karakula-Juchnowicz and Stecka, 2017) employed common words of foreign origin, that can't be read correctly using Polish orthographic principle (e.g., *popcorn*, *rock*, or *déjà vu*). However, it must be acknowledged that the primary purpose of such instruments is to measure crystallized intelligence in people affected with dementia, rather than print exposure *per se*.

Word decoding and recognition

In the narrow sense, decoding is defined as the ability to apply the working knowledge of grapheme-phoneme correspondence rules to translate print into sound (phonology). As such, decoding is usually operationalized through accuracy and speed or pronouncing printed words or pseudowords. In the broader sense, decoding may be understood as the ability to access the phonology of any words, whether familiar or unfamiliar ones, without having to rely on contextual cues—, i.e., it encompasses context-free word recognition. As reported in the metaanalysis including 13 studies of 2,440 adult English speakers with low reading skills, correlation of decoding and reading comprehension in this group is significant and moderate (avg. r = 0.52; Tighe and Schatschneider, 2016).

Poor decoding (despite adequate instruction) is also the hallmark of dyslexia (Rose, 2009). It is often related to other cognitive or metalinguistic factors playing role in dyslexia, such as phonological awareness, rapid auditory processing, visual processing, or automatization (Shaywitz and Shaywitz, 2005). Some adults with low functional literacy skills meet criteria of dyslexia (Vágvölgyi et al., 2021), though the extent to which cognitive deficits typical of dyslexia are the root cause of low functional literacy remains unclear. In any case, the assessment of decoding skills should be included in the assessment of functional literacy—weak decoding skills do explain why some readers fail to understand, or engage with, more complex texts. Functional literacy skills are unlikely to develop unless decoding problems are remedied or at least compensated for.

Decoding can be operationalized with various single word or pseudoword reading, word-to-picture matching, or lexical decision tasks. The method of measurement used should be taken into consideration; another metaanalysis of English studies showed that the relationship of reading comprehension and decoding can vary from non-significant r = 0.39 in the lexical decision task to r = 0.86 in accuracy of word reading across all age grups (García and Cain, 2014). In the English language literature, decoding is usually measured with standard tests, e.g., Test of Word Reading Efficiency (TOWRE), Word Reading test in the Wide Range Achievement Test (WRAT) or Word Attack in the Woodcock-Johnson (WJ IV). The majority of reading performance tests were developed for the assessment in the context of education, though adult norms are sometimes available (e.g., norms for TOWRE extend up to 24 years of age, while WRAT and WJ IV Test of Achievement norms cover all age ranges).

Tasks of this kind are sometimes criticized as lacking ecological validity—after all, meaningful engagement of print does not involve sounding out unfamiliar (let alone meaningless) words out of context. However, this critique misses the point. Measures of decoding are valid theoretically insofar as they measure essential cognitive components of the reading process. They also have predictive validity in that they correlate with more ecologically valid reading outcomes. However, it must be acknowledged that such methods may favor better readers, for whom reading aloud is less stressful (McLaughlin, 1997).

One way to overcome the problem of reading aloud might be to use a silent reading task. One of the options is a lexical decision task, in which participants have to decide as quickly and accurately as possible whether a letter string is a real word or not. It can be done with the paper-pencil version or a computerized task. A recent example is the Rapid Online Reading Assessment (ROAR), designed for children as well adults. Reaction times in ROAR turned out to be highly reliable and correlated with the WJ IV Letter Word Identification test, which is one of the most widely used standardized measures of decoding (Yeatman et al., 2021). The tool is in the public domain (https://roar.stanford.edu/), the code is shared on GitHub and available for other languages' adaptations. The go/no-go procedure as an alternative to the yes/no lexical decision task was also proposed in the study including university students (Perea et al., 2002). In PIAAC, the functional literacy assessment is augmented by a "reading component" skills test aimed at garnering information regarding adults with limited literacy skills. This encompasses a basic set of decoding skills essential for individuals to derive meaning from written texts, including simple vocabulary knowledge and decoding. They use a word to picture matching paradigm, where the respondent has to circle the printed word (out of four) that corresponds to the picture (Pellizzari and Fichen, 2017).

It is important to note that word recognition is related to vocabulary knowledge, and vocabulary, on the other hand, is closely related to reading comprehension (Nation, 2009). Again, most studies in this area involved children, and available tests are also mostly designed for, and normed on, children. However, the relationship between vocabulary, phonology, decoding, and reading comprehension in children and adults with low literacy skills is not identical. For example, it was shown that children and adults tested with the Peabody Picture Vocabulary Test—IIIB had different response patterns. Relative to children, adults scored lower on easier items and higher on more difficult items (Pae et al., 2012).

Another skill closely related to decoding is phonological awareness. The Comprehensive Test of Phonological Processing (CTOPP) is widely used in the English language context, even though the norms do not extend beyond 24 years of age. Still, CTOPP is commonly used by adult literacy researchers, and also recommended as a tool suitable for adult literacy educators (Nanda et al., 2014). However, a study controlling the psychometric attributes of CTOPP in adults with lower literacy skills found that its reliability and validity are very limited and results should be treated with caution (Nanda et al., 2014). Languages other than English may face similar scarcity of the measures of decoding (and related cognitive skills) that are valid in adult populations. Moreover, the relationship between decoding, comprehension and other reading-related skills may vary cross-linguistically, especially between alphabetic and non-alphabetic writing systems (Share, 2021). To sum up, we believe that norms developed for decoding and other reading-related skills, if to be used with adults on lower literacy spectrum, should include individuals of all age groups, but also varying education and professional backgrounds. The level of decoding skill indispensable for functional reading should be established, with cut-off points demarcating high risk of functional reading problems.

Reading comprehension

Sentence reading tests and fill-in-the-blank reading tests

Sentence reading and close reading tasks may be considered more ecologically valid, even if less theoretically pure, measures of decoding, more similar to "natural" reading situations. In these tasks, whole meaningful sentences, rather than disconnected lists of words or pseudowords, need to be processed. Here, decoding skills are inevitably confounded with comprehension—but it is a "local" comprehension, which does not demand inferences or evaluations typical for passage reading tests.

The authors of the ROAR test also proposed the Sentence Reading Efficiency test (ROAR-SRE), which is currently under development and also available on the website (https://roar. stanford.edu/). This involves silent reading of short statements and deciding, for each statement, whether it is true or false. The team explored the possibility of the automated generation of items for the sentence reading test using Chat GPT3. They found that generated items closely resemble standardized test items in terms of their factual ambiguity, content appropriateness, and complexity (White et al., 2022).

In a fill-in-the-blank test (cloze test), a participant is asked to write or select a word in short text passages. Such tests were criticized as measuring only "local" comprehension, related to lexical, syntactic, and grammatical awareness (Carlisle and Rice, 2004). However, there is evidence that the score on a fill-in-the-blank test highly correlates with the long reading comprehension test scores, especially in the lower literacy population (Gellert and Elbro, 2013).

In PIAAC, sentence processing is measured in the reading components module through tasks designed to evaluate a respondent's ability to understand the logical coherence and sensibility of sentences in real-world contexts. Participants might be asked to silently read sentences like "Three girls ate the song" or "The man drove the green car" and judge their logical validity. Such tasks gauge an individual's capacity to process linguistic information, discern meaning, and evaluate the plausibility of given textual scenarios. PIAAC reading components also measure passage comprehension using a variant of cloze procedure, whereby respondents fill the gap by choosing the suitable word from the alternative provided. For instance, in a letter addressing the rise in bus fare, participants might encounter sentences such as, The price will go up by twenty percent starting next wife/month. Participants would need to determine that month is the correct choice (Sabatini and Bruce, 2009).

Reading single words requires the processing of orthographic and phonological information—and also semantic information if the task so requires (e.g., word to picture matching). Comprehension of sentences must also involve our syntactic competencies. But to fully comprehend what we read, more has to be done: integrating the meaning of multiple sentences, connecting to the background knowledge, generating inferences, identifying the text structure, and considering the authors' goals and motives (Graesser et al., 1994). It is reflected by Graesser and McNamara's theoretical framework identifying six levels of comprehension: (1) *word*; (2) *syntax*; (3) the explicit *text base* related to the literal content of the text; (4) the referential *situation model* including inferences activated by the explicit text; (5) the *genre/rhetorical*

structure focusing on the category of text and its composition; (6) and the *pragmatic communication level,* involving context-sensitive exchanges between reader and text. Only the first two levels represent the basic reading components; the others are related to higher-level, more complex semantic and discourse processing. From a broader perspective, comprehension depends on linguistic (vocabulary, syntax, etc.) as well as nonlinguistic (e.g., monitoring, working memory) competencies (Kendeou et al., 2016)—but also relevant knowledge and experience. Motivation, attention, background knowledge, and context familiarity—all need to be on an adequate level for successful text comprehension. A failure to control for individual differences in motivation or background knowledge is a weakness of assessments (see: Sabatini et al., 2013). For example, adults who left formal education early find it difficult to process more complex syntactic structure which occur mostly in writing and not in speech (Dąbrowska, 1997); if the test uses such language structures it won't be valid for the test taker (Snyder et al., 2005). Many reading comprehension tests use expository or literary texts, similarly to school textbooks. In the context of educational assessment, they are valid, as they sample uniformed texts that students deal with on an everyday basis. While such school-like tests don't measure adult functional literacy per se, the processes involved are important as well during the lecture on functional reading materials. As such, if no functional tests are available, such non-ecological tests provide some proxy for functional reading skills.

The goal of this section isn't to review all possible reading comprehension tests for adults; those available online (sometimes as demos), documented, or published as peer-reviewed articles are almost exclusively in English (Lan and Yu, 2022). Rather, we are trying to describe general trends in adult reading comprehension assessment and pinpoint some issues with them.

Types of comprehension

The current PIAAC study (cycle 2) defines three broad cognitive strategies used when responding to written texts: (1) accessing, (2) understanding, and (3) evaluating/reflecting. Two types of accessing the information are distinguished: selecting the relevant text (i.e., the one containing the required information) in a set of texts, and locating information within that text. Understanding is divided into literal comprehension, inferential comprehension, and multiple-text inferential comprehension when information needs to be integrated across two or more texts. Finally, competent readers are expected to be able to critically assess the quality of information in a text. Evaluation is important for the selection of best sources, and protection against misinformation and propaganda. Evaluation can be based on assessing the accuracy, soundness, and task relevance of a text (OECD, 2021). Not all three operations are expected to be available for everybody—tasks adequate for adults at PIAAC level 1 or below only involve identifying literal information in the short texts. Evaluating and reflecting are the skills involved in tasks with higher demand.

Other tests of reading comprehension may be constructed around somewhat different distinctions, e.g., between literal, inferential and evaluative comprehension, or between memory (where literal repetition of information from the passage is

required) factual and inferential comprehension (Brooks et al., 2016). In yet other tests, the comprehension processes are divided solely into "local" (related to single sentences) and "global" (associated with the whole text). Nonetheless, quite often these "subprocesses" of comprehension are neither clearly defined nor rooted in theoretical models (Sabatini et al., 2013).

Text types

PIAAC assessment includes both traditional "running text" composed of sentences and paragraphs, as well as digital texts containing interactive navigation tools (such as tables of contents, diagrams, or hyperlinks) or images that server as functional rather than decorative elements. Such interactive texts are sometimes referred to as "discontinuous" or "noncontinuous" texts. They seem more ecologically valid; outside the educational contexts and reading books for pleasure, texts that adults interact with, especially online, are typically accompanied by other textual or graphic elements. However, the German NEPS made a different decision and based their assessment solely on continuous text passages (Gehrer et al., 2013). They argue that continuous texts are still the primary text type, and reading continuous and discontinuous texts requires different types of comprehension processes, i.e., integrating the mental representation of the text and the images. Nonetheless, the authors admit that their approach is less ecologically valid and only partially corresponds to what reading means in modern everyday life. Most standardized tests of reading comprehension used for clinical and research purposes also rely solely on continuous texts and omits digital texts and multimedia, which some consider a limitation (Sabatini et al., 2013). On the other hand, if the test focuses on functional literacy and identification of low-skilled adults, it is more likely touse shorter, discontinuous, and digital tests: e.g., Level-One Study (Buddeberg et al., 2020), Literacy-Assessment and Monitoring Programme (Ercikan et al., 2008), Comprehensive Adult Student Assessment System (Gorman and Ernst, 2004), and Test of Adult Basic Education (TABE).

Another issue is the *topic* of a text. Narrative and expository texts are used most often in the assessment of children and adolescents, and also in some tests designed for adults (e.g., Adult Reading Test, Brooks et al., 2016). However, if the goal is to measure adult functional literacy, then texts related to everyday life are typically used. The current cycle of PIAAC divides reading materials into (1) personal, (2) work and occupation, and (3) societal/community-related (Rouet et al., 2021). Personal-oriented texts can be, related to personal finances, housing, insurance, interpersonal relationships, health, and safety issues (e.g., disease treatment, first aid, and prevention), consumer habits (e.g., credit and banking, advertising, or making purchases). They can also include leisure and recreation (e.g., traveling, eating out, and gaming). Work and occupation-related texts focus on finding employment, finances, and handling the job (e.g., regulations, organization, safety instructions), but at the same time avoid very job-specific texts, which would pose the problem of background knowledge. Finally, social and civic contexts use texts related to dealing with community resources, public services, and staying informed, but also opportunities for further learning (Rouet et al., 2021).

Answer formats

There are a few options on how to collect participants' responses in the reading comprehension task: multiple choice questions, or shorter and longer open-ended questions, either general or more specific (wh_ questions). In cloze tasks, a blank has to be filled in or selected out of a few options. Some other variants are possible; e.g., in a MOCCA test, respondents have to select the correct final sentence for the passage they have read, which may be considered the higher-level version of the fill-in-the blank test. In the Literacy Level of the CASAS Life Skills Reading assessment, adult students do not have to enter their answers on a separate test record; they can mark their answers directly in the test booklet.

The over-reliance on multiple-choice format in the reading comprehension tests was criticized as emphasizing strategic reasoning over the understanding of the text (Sabatini et al., 2013). Multiple choice answers provide cues as to how to approach the task, e.g., what target information needs to be scanned for; and, consequently, they influence the "natural" reading process (Rupp et al., 2006). Moreover, respondents who are familiar with the multiple choice format will approach the test differently compared to those who lack such familiarity. Ozuru et al. (2013) compared university students' answers in open-ended and multiple-choice tests and found that good answers to multiple-choice questions were more related to the prior knowledge of the subject. Finally, multiple-choice comprehension questions are strongly related to decoding skills, as there is simply more to read, i.e., both questions, answers and foils (Cain and Oakhill, 2006; Rouet et al., 2021). On the other hand, answering open-ended questions is more time-consuming and produces missing data (Reja et al., 2003), and shows another type of bias: men, younger participants, and those with more digital experience in web surveys answered more often than others in the online tasks (Zhou et al., 2017). Moreover, written responses tap on not only comprehension skills but also written production (Rouet et al., 2021). Importantly, scoring open-ended answers requires subjective judgment and so faces the problem of imperfect inter-rater reliability—not to mention it is much more time-consuming. In general, the researcher needs to consider how the response format might affect the performance of different groups and confound comprehension with other, related skills (Rouet et al., 2021).

Dependence problem

Some comprehension tests can be responded to at the above-chance level by using only the background knowledge, without reading the text at all. This issue was first recognized more than 50 years ago (Pyrczak, 1972; Tuinman, 1973). Passage Dependency Index (PDI) was proposed to estimate which test items really require reading the relevant passage for correct response, and which not (Tuinman, 1973) noted that many common tests lack this valid dependence, i.e., can be responded to correctly based on background knowledge, without recourse to the information conveyed in the text. Even patients with aphasia were shown to give above chance answers to multiple choice reading comprehension questions, without reading the relevant passages (Nicholas et al., 1986). Unfortunately, it appears that the problem still holds in the 21st century, even for some of the most commonly used reading comprehension tests: The Nelson-Denny Reading Test (NDRT) and Gray Oral Reading Test (GORT) (Stevens and Price,

1999). GORT and NDRT were used in 7 out of all 14 studies on adult comprehension reported in the meta-analysis (García and Cain, 2014). These two tests should probably not be used in adult assessment at all; both are normed for adults up to 24 years of age only. Greenberg et al. (2009) showed that GORT exhibits odd properties when administered to adult poor readers: its comprehension subtest scores failed to correlate with its accuracy, rate and fluency subtest scores, and showed only weak correlation with other (non-GORT) measures of reading and reading-related skills. These patterns are quite different from those reported for children, and question the validity of GORT with adult samples. Nonetheless, given the widespread use of GORT and NDRT, and also other tests relying on common-sense questions, it is important to highlight the issue of the passage independence of items.

Keenan and Betjemann (2006) showed that undergraduates answered 86% of GORT questions were answered with above-chance accuracy without reading the relevant text; the pattern of responses was similar in children, who had slightly lower scores. Similarly, Coleman et al. (2010) showed that university students answered 70–80% of NDRT multiple choice questions above chance without reading the relevant text. This was the case for both literal and inference questions. Accuracy rates were exceptionally high for science questions, whose content has been adapted from high school textbooks. Students at risk of learning difficulties were almost as good as the others. Still, we can speculate that how easy it is for a person to solve a passageless test may depend on their academic background—a potentially serious bias. Another risk of using tests burdened with passage dependence problems is the probability of overestimating reading comprehension skills and false-negative diagnoses (Coleman et al., 2010). On the other hand, even if general knowledge usually facilitates reading comprehension—people knowledgeable about a given topic are better, faster, and more accurate than novices—inaccurate background knowledge can also induce incorrect inferences (Kendeou et al., 2016).

The passage dependence issue should be alarming for the test makers. A good idea would be to check the questions' independence upon test construction (Tuinman, 1973). Within the test, we recommend the use of entirely fictional texts (Lifson et al., 1984) or ecological, everyday life reading materials with naturally high variability, where the correct information cannot be inferred from prior experience. An example of such text would be a lost dog poster (When and where it went missing? How to contact the owner?). Such functional texts are known to everybody from everyday life, they serve specific functions, and often are useful for adult readers (Napitupulu and Napitupulu, 2020). Being more relevant, they might be more interesting and keep reading motivation higher—especially for those with lower literacy skills. Our suggestion is to include such tests in the assessment of functional reading skills—just as the PIAAC study does.

New assessment ideas in managing the test-creator bias in adult literacy assessment

The functionality of literacy in childhood is mainly based on relatively homogenous curricular requirements, designed by the educational program creators, and therefore easy to operationalize. One of the greatest challenges in the assessment of adult reading comprehension is the diversity of spontaneously emerging functionality of literacy in various contexts in which adults function. Until recently, the sampling of the universe of functional literacy for the purpose of assessment had to rather arbitrary. However, recent advances in machine learning enabled the creation of more objective assessments. On one hand, assessment material can now be based on the selection of knowledge that adults really find appealing; on the other hand, it can be based on samples taken from a much larger text corpora. Two of such projects will be described below.

The Boolean Questions (BoolQ) literacy assessment (Clark et al., 2019) defines its literacy area by focusing on naturally occurring questions from Google queries and Wikipedia articles corresponding to such questions. The questions are not prompted by researchers, and therefore the topics that they relate to are sampled from a natural repertoire of human curiosity, which emerges in spontaneous contexts. The core of the literacy assessment in BoolQ lies in the ability to correctly answer a yes/no (boolean) question based on a passage from a Wikipedia article. The subjective element lies in the fact that the Wikipedia passages have been previously marked by a group of human annotators as containing relevant information sufficient for answering the question (Kwiatkowski et al., 2019). Nonetheless, the content of this assessment is not biased by the subjective, test-creator's notions of what is, or what is not, worthy of comprehending. Furthermore, people who ask questions via Google queries do not even know whether they are answerable. This protects this type of assessment from a common "school knowledge" bias, where the literacy assessment refers to those areas of knowledge where the test-creator has high certainty about correct answers. Interestingly, it appears that people, when asking natural questions, do not seek knowledge verification, but form their queries in a more open-ended, curiosity-driven way (Clark et al., 2019). For example, people prefer to ask "Has the UK ever been hit by a hurricane?" rather than "Has the UK been hit by a hurricane in 1905?". This seems understandable and also stands in contrast with a typical test-taking situation where the test-designer asks questions they already know answers to.

When it comes to the complexity of comprehension skills involved, even though the final answer to a question in BoolQ is simply a boolean (which can only have two values: true or false), it requires making different types of inferences. Results obtained by researchers show that simple paraphrasing from Wikipedia is enough to detect no more than 40% of the answers to the Google query questions (Clark et al., 2019). The majority of answers require additional, more complex reasoning, such as reaching a conclusion based on what is missing in the Wikipedia article or using general world knowledge to connect statements in the question and the passage. Generally, this type of assessment nicely captures the functionality of the adult lifelong learning process in a society with internet access and proficiency in English as a primary/secondary language.

Another example of a novel literacy assessment method, the Reading Comprehension with Commonsense Reasoning Dataset (ReCoRD) is based on an impressive sample of 120,000 passages from news articles (Zhang et al., 2019). These news articles were automatically downloaded from popular news sources such as CNN and the Daily Mail. Just as with BoolQ, ReCoRD offers

comprehension test items that are not subject to the elicitation bias. ReCoRD additionally goes beyond boolean questions and includes fill-in-the-blank comprehension questions, which are also automatically generated. The generation procedure utilizes the typical structure of a news article. In this structure, a summary is followed by the full article and supplemented by bullet points, therefore some forms of data compression and redundancy are already present and can be extracted for testing purposes. For example, the redundant parts of the message can be used for the fill-in-the-blank questionnaire items.

There is an interesting side note. Since the authors of ReCoRD wanted to compare machine learning algorithms, with the common-sense literacy performance of, what is labeled as a "typical human reader", they included popular press news items and fill-in-the-blank questions, for which the "human-level performance" reached about 90% correct rate in their study. The trouble is that we don't yet know what an average human literacy performance is in this context. In ReCoRD, as in many other similar research programs, the "human-level performance" indicator is composed of results of a very specific demographic, predominantly users of online work platforms such as Amazon Mturk or hybrid.io (Clark et al., 2019). In this particular case, the pool of Amazon Mturk users was further narrowed down to only the most reliable individuals with previous experience in text-based assignments (95% acceptance rate, a minimum of 50 previously completed assignments), located in the US, Canada, or Great Britain (Zhang et al., 2019). Furthermore, workers in this study were blocked, and their work reassigned, if their average accuracy was lower than 20%. Therefore, the level of text comprehension in AI studies, casually labeled as a "commonsense level of human literacy" (Zhang et al., 2019) is most likely a gross overestimation coming from a very WEIRD sample (Henrich et al., 2010). If this bias persists, and we do not engage in better ways to estimate the levels of adult human literacy, we will greatly underestimate the true impact of recent technological advances in machine learning.

To summarize, machine learning text compression attempts can be extremely useful for designing adult text comprehension measurement because they are based on relatively unbiased and naturally occurring text prompts. The thematic selection of texts is not based on the test-maker notion of importance, but rather on bottom-up functionality, and the sampling of test materials is extremely broad and covers a wide range of topics and contexts.

Closing remarks

In this paper, we focused on a functional level of literacy; a kind of literacy that enables a reader to decode public transportation schedules and shopping receipts; necessary skills for dealing with daily life tasks. But there are ways of interacting with a text beyond word decoding and basic semantic retrieval, and there are texts that go beyond short factual statements (Schüller-Zwierlein et al., 2022). Immersive, critical reading of longer, literary forms is related, among all, to better concentration, perspective taking, imagination, and empathy; it enhances understanding of intertextuality and contextuality, metaphorical expressions, and social interactions (e.g., Jerrim and Moss, 2019; Wicht et al., 2021). Savoring of a literary text demands repeated comprehension

monitoring, reflection over a sustained time, and discovering several layers of meaning, which develops deeper, metacognitive comprehension levels (Lacy, 2014). Nonetheless, we still lack theoretical perspectives and empirical tools to measure complex literary reading behaviors, lack empirical data, and, finally, long-term reading education strategies, acknowledging the role of higher-level reading skills in society (Schüller-Zwierlein et al., 2022).

What should receive more attention are the predictors of performance on functional literacy measures—both to diagnose the source of the possible problems and propose a sensible support or intervention promoting reading related skills. A recent study found that having breakfast was a significant predictor of PIRLS scores in Nordic countries, even after controlling for SES (Illøkken et al., 2022). Sleep deprivation also impacts reading, language, and cognitive abilities, influencing test performance, but also promoting the tendency to skip the instructions (Mathew et al., 2018). Hearing and vision deficits may have gone uncorrected in some adults and also should be checked upon the diagnosis, along with cognitive skills such as rapid automatized naming, working memory, and attention (Sabatini et al., 2019). Such factors should not be overlooked in the literacy assessment, both in research and practical contexts. However, controlling known cognitive factors related to reading (like the phonology, rapid automatized naming levels, and others) explains only around 30% of the variance in reading (e.g., Compton et al., 2001; Debska et al., 2021). Models of reading developed when researching children are inadequate for adults (Sabatini et al., 2019). This indicates that studies of reading should consider more context-dependent, socioeconomic, and literacy practices factors when explaining the reading level. Another crucially important issue is motivation, engagement, and self-efficacy in reading—important both in the context of assessment and everyday life literacy practices.

Conclusion

In Europe, virtually all people are literate—but still around 15% of Europeans assessed in the first cycle of PIAAC are only literate below the "functional" level (OECD, 2012), i.e., they may encounter problems in their everyday functioning when reading is demanded. This fact alone shows that we need valid tools to evaluate adult functional literacy difficulties, as well as the causes of these difficulties, whether constitutional (e.g., cognitive deficits resulting in poor decoding skills) or environmental.

Existing research on literacy is predominantly focused on children. Literacy assessment is also mainly targeted at children, and there is a need for tests developed specifically for adult populations—especially low skills adult populations. Such tests should focus on functional literacy, implying a shift away from academic content. In developing these tests, emphasis should be placed on everyday life experiences and interactions. Recent developments in AI text comprehension can offer valuable insights. Queries people enter into search engines, content of Wikipedia pages, and news articles from the popular press can serve as the backbone of a naturalistic, functional assessment of adult reading related skills. Automated item creation or assessment could be valuable as well. These new methods could be supplemented

with measurement of basic cognitive processes, such as decoding abilities, which are already covered adequately by existing literature.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Acheson, D. J., Wells, J. B., and MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behav. Res. Methods* 40, 278–289. doi: 10.3758/BRM.40.1.278

Applegate, A. J., Applegate, M. D., Mercantini, M. A., McGeehan, C. M., Cobb, J. B., DeBoy, J. R., et al. (2014). The Peter effect revisited: reading habits and attitudes of college students. *Literacy Res. Instruct.* 53, 188–204. doi: 10.1080/19388071.2014.898719

Atcherson, S. R., Richburg, C. M., Zraick, R. I., and George, C. M. (2013). Readability of questionnaires assessing listening difficulties associated with (central) auditory processing disorders. *Lang. Speech Hear. Serv. Sch.* 44, 48–60. doi: 10.1044/0161-1461(2012/11-0055)

Baker, D. W., Williams, M. V., Parker, R. M., Gazmararian, J. A., and Nurss, J. (1999). Development of a brief test to measure functional health literacy. *Patient Educ. Couns.* 38, 33–42. doi: 10.1016/S0738-3991(98)00116-5

Barnes, A. E., and Kim, Y. S. (2016). Low-skilled adult readers look like typically developing child readers: a comparison of reading skills and eye movement behavior. *Read. Writ.* 29, 1889–1914. doi: 10.1007/s11145-016-9657-5

Boudard, E., and Jones, S. (2003). The IALS approach to defining and measuring literacy skills. *Int. J. Educ. Res.* 39, 191–204. doi: 10.1016/j.ijer.2004.04.003

Bourdieu, P. (1991). *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.

Brice, J. H., Foster, M. B., Principe, S., Moss, C., Shofer, F. S., Falk, R. J., et al. (2014). Single-item or two-item literacy screener to predict the s-tofhla among adult hemodialysis patients. *Patient Educ. Couns.* 94, 71–75. doi: 10.1016/j.pec.2013.09.020

Brooks, P., Everatt, J., and Fidler, R. (2016). *Adult Reading Test 2 (Silent and Oral Reading) A UK Standardised Test for Prose Reading Accuracy, Comprehension and Speed, With Norms for Writing Speeds*. Adult Reading Test Limited.

Brysbaert, M., Sui, L., Dirix, N., and Hintz, F. (2020). Dutch author recognition test. *J. Cogn.* 3:6. doi: 10.5334/joc.95

Buddeberg, K., Dutz, G., Grotlüschen, A., Heilmann, L., and Stammer, C. (2020). Low literacy in germany: results from the second german literacy survey. *Eur. J. Res. Educ. Learn. Adults* 11, 127–143. doi: 10.3384/rela.2000-7426.rela9147

Cain, K., and Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *Br. J. Educ. Psychol.* 76, 683–696. doi: 10.1348/000709905X67610

Carlisle, J., and Rice, M. (2004). "Assessment of reading comprehension," in *Handbook of Language and Literacy*, 521–555.

Castell, S., de, Luke, A., and MacLennan, D. (1981). On defining literacy. *Can. J. Educ./Revue Canadienne de l'éducation* 6, 7–18. doi: 10.2307/1494652

Chen, S. Y., and Fang, S. P. (2015). Developing a Chinese version of an Author Recognition Test for college students in Taiwan. *J. Res. Read.* 38, 344–360. doi: 10.1111/1467-9817.12018

Clark, K., Luong, M. T., Khandelwal, U., Manning, C. D., and Le, Q. V. (2019). Bam! born-again multi-task networks for natural language understanding. *arXiv* preprint arXiv:1907.04829. doi: 10.48550/arXiv.1907.04829

Coleman, C., Lindstrom, J., Nelson, J., Lindstrom, W., and Gregg, K. N. (2010). Passageless comprehension on the nelson-denny reading test: well above chance for university students. *J. Learn. Disabil.* 43, 244–249. doi: 10.1177/0022219409345017

Compton, D. L., Defries, J. C., and Olson, R. K. (2001). Are RAN- and phonological awareness-deficits additive in children with reading disabilities? *Dyslexia* 7, 125–149. doi: 10.1002/dys.198

Corredor, C. M., Ferreira, S. C. O., Monsalve, A. M. S., and Currea, A. M. (2019). "Piloting a self-report questionnaire to detect reading-writing difficulties in students from two Colombian universities," in *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT), Vol. 2161* (IEEE), 311–313.

Dąbrowska, E. (1997). *The LAD Goes to School: A Cautionary Tale for Nativists*. Berlin: de Gruyter.

De Greef, M., van Deursen, A. J. A. M., and Tubbing, M. (2013). Development of the DIS-scale (Diagnostic Illiteracy Scale) in order to reveal illiteracy among adults. *J. Study Adult Educ. Learn.* 1, 37−48.

Debska, A., Banfi, C., Chyl, K., Dziegiel-Fivet, G., Kacprzak, A., Łuniewska, M., et al. (2021). Neural patterns of word processing differ in children with dyslexia and isolated spelling deficit. *Brain Struct. Funct.* 226, 1467–1478. doi: 10.1007/s00429-021-02255-2

Ercikan, K., Arim, R., Oliveri, M., and Sandilands, D. (2008). *Evaluation of the Literacy Assessment and Monitoring Programme (LAMP)/UNESCO Institute for Statistics (UIS)*. UNESCO Document IOS/EVS/PI/91.

Feng, L., Hancock, R., Watson, C., Bogley, R., Miller, Z. A., Gorno-Tempini, M. L., et al. (2022). Development of an Abbreviated Adult Reading History Questionnaire (ARHQ-Brief) using a machine learning approach. *J. Learn. Disabil.* 55, 427–442. doi: 10.1177/00222194211047631

Frijters, J. C., Brown, E., and Greenberg, D. (2019). "Gender differences in the reading motivation of adults with low literacy skills," in *The Wiley Handbook of Adult Literacy*, 63–87. doi: 10.1002/9781119261407.ch3

García, J. R., and Cain, K. (2014). Decoding and reading comprehension: a meta-analysis to identify which reader and assessment characteristics influence

the strength of the relationship in English. *Rev. Educ. Res.* 84, 74–111. doi: 10.3102/0034654313499616

Gehrer, K., Zimmermann, S., Artelt, C., and Weinert, S. (2013). *NEPS Framework for Assessing Reading Competence and Results from an Adult Pilot Study.* doi: 10.25656/01:8424

Gellert, A. S., and Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *J. Psychoeduc. Assessm.* 31, 16–28. doi: 10.1177/0734282912451971

Gilger, J. W. (1992). Using self-report and parental- report survey data to assess past and present academic achievement of adults and children. *J. Appl. Dev. Psychol.* 13, 235–256. doi: 10.1016/0193-3973(92)90031-C

Giménez, A., Luque, J. L., López-Zamora, M., and Fernández-Navas, M. (2015). A self-report questionnaire on reading-writing difficulties for adults [Autoinforme de Trastornos Lectores para AdultoS (ATLAS)]. *Anales de Psicología/Ann. Psychol.* 31, 109–119. doi: 10.6018/analesps.31.1.166671

Gorman, D., and Ernst, M. L. (2004). Test review: the comprehensive adult student assessment system (CASAS) life skills reading tests. *Lang. Assess. Q.* 1, 73–84. doi: 10.1207/s15434311laq0101_8

Graesser, A. C., Singer, M., and Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychol. Rev.* 101, 371–395. doi: 10.1037/0033-295X.101.3.371

Greenberg, D., Pae, H. K., Morris, R. D., Calhoon, M. B., and Nanda, A. O. (2009). Measuring adult literacy students' reading skills using the Gray Oral Reading Test. *Ann. Dyslexia* 59, 133–149. doi: 10.1007/s11881-009-0027-8

Hamilton, M., and Barton, D. (2000). The international adult literacy survey: what does it really measure? *International Review of Education* 46, 377–389. doi: 10.1023/A:1004125413660

Hautecoeur, J.-P. (2000). Editorial introduction: literacy in the age of information: knowledge, power or domination? *Int. Rev. Educ./Internationale Zeitschrift Für Erziehungswissenschaft/Revue Internationale de l'Education* 46, 357–365. doi: 10.1023/A:1004129812751

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world?. *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X

Hoover, W. A., and Gough, P. B. (1990). The simple view of reading. *Read. Writ.* 2, 127–160. doi: 10.1007/BF00401799

Illøkken, K. E., Ruge, D., LeBlanc, M., Øverby, N. C., and Vik, F. N. (2022). Associations between having breakfast and reading literacy achiev2ement among nordic primary school students. *Educ. Inqu.* 20, 1–13. doi: 10.1080/20004508.2022.2092978

Jerrim, J., and Moss, G. (2019). The link between fiction and teenagers' reading skills: International evidence from the OECD PISA study. *Br. Educ. Res. J.* 45, 181–200. doi: 10.1002/berj.3498

Karakula-Juchnowicz, H., and Stecka, M. (2017). Polish Adult Reading Test (PART)-construction of Polish test for estimating the level of premorbid intelligence in schizophrenia. *Psychiatr. Pol.* 51, 673–685. doi: 10.12740/PP/OnlineFirst/63207

Keenan, J. M., and Betjemann, R. S. (2006). Comprehending the gray oral reading test without reading it: why comprehension tests should not include passage-independent items. *Scientif. Stud. Read.* 10, 363–380. doi: 10.1207/s1532799xssr1004_2

Kendeou, P., McMaster, K. L., and Christ, T. J. (2016). Reading comprehension: core components and processes. *Policy Insights Behav. Brain Sci.* 3, 62–69. doi: 10.1177/2372732215624707

Kirsch, I., and Guthrie, J. T. (1977). The concept and measurement of functional literacy. *Read. Res. Q.* 485–507. doi: 10.2307/747509

Kirsch, I., and Lennon, M. (2017). PIAAC: a new design for a new era. *Large-Scale Assessm. Educ.* 5, 11. doi: 10.1186/s40536-017-0046-6

Krashen, S., and Kim, H. (1998). The Author Recognition Test without foils as a predictor of vocabulary and cultural literacy test scores. *Percept. Mot. Skills* 87, 544–546. doi: 10.2466/pms.1998.87.2.544

Kwiatkowski, T., Palomaki, J., Redfield, O., Collins, M., Parikh, A., Alberti, C., et al. (2019). Natural questions: a benchmark for question answering research. *Trans. Assoc. Comput. Linguist.* 7, 452–466. doi: 10.1162/tacl_a_00276

Lacy, M. (ed.). (2014). *The Slow Book Revolution: Creating a New Culture of Reading on College Campuses and Beyond.* Bloomsbury Publishing.

Lan, X., and Yu, Z. (2022). A bibliometric review study on reading literacy over fourteen years. *Educ. Sci.* 13:27. doi: 10.3390/educsci13010027

Lee, H., Seong, E., Choi, W., and Lowder, M. W. (2019). Development and assessment of the Korean Author Recognition Test. *Quart. J. Experim. Psychol.* 72, 1837–1846. doi: 10.1177/1747021818814461

Lefly, D. L., and Pennington, B. F. (2000). Reliability and validity of the adult reading history questionnaire. *J. Learn. Disabil.* 33, 286–296. doi: 10.1177/002221940003300306

Lifson, S., Scruggs, T. E., and Bennion, K. (1984). passage independence in reading achievement tests: a follow-up. *Percept. Mot. Skills* 58, 945–946. doi: 10.2466/pms.1984.58.3.945

Mar, R. A., and Rain, M. (2015). Narrative fiction and expository nonfiction differentially predict verbal ability. *Scientif. Stud. Read.* 19, 419–433. doi: 10.1080/10888438.2015.1069296

Martin, A., Schurz, M., Kronbichler, M., and Richlan, F. (2015). Reading in the brain of children and adults: a meta-analysis of 40 functional magnetic resonance imaging studies. *Hum. Brain Mapp.* 36, 1963–1981. doi: 10.1002/hbm.22749

Mathew, G. M., Martinova, A., Armstrong, F., and Konstantinov, V. (2018). The role of sleep deprivation and fatigue in the perception of task difficulty and use of heuristics. *Sleep Sci.* 11, 16. doi: 10.5935/1984-0063.20180016

McCarron, S. P., and Kuperman, V. (2021). Is the author recognition test a useful metric for native and non-native English speakers? An item response theory analysis. *Behav. Res. Methods* 53, 2226–2237. doi: 10.3758/s13428-021-01556-y

McLaughlin, M. (1997). Basic Writers Three Years Later: Their Problems and Their Priorities. Published: Distributed by ERIC Clearinghouse, 1997.

Möller, J., and Bonerad, E. M. (2007). "Habitual reading motivation questionnaire," in *Psychologie in Erziehung und Unterricht.* doi: 10.1037/t58194-000

Moore, M., and Gordon, P. C. (2015). Reading ability and print exposure: item response theory analysis of the author recognition test. *Behav. Res. Methods* 47, 1095–1109. doi: 10.3758/s13428-014-0534-3

Morris, N. S., MacLean, C. D., Chew, L. D., and Littenberg, B. (2006). The single item literacy screener: evaluation of a brief instrument to identify limited reading ability. *BMC Fam. Pract.* 7, 21. doi: 10.1186/1471-2296-7-21

Murray, T. S. (2017). *Functional Literacy and Numeracy: Definitions and Options for Measurement for the SDG Target 4.6.* UNESCO.

Nanda, A. O., Greenberg, D., and Morris, R. D. (2014). Reliability and validity of the CTOPP Elision and Blending Words subtests for struggling adult readers. *Read. Writ.* 27, 1603–1618. doi: 10.1007/s11145-014-9509-0

Napitupulu, S., and Napitupulu, F. D. (2020). *The Functional Texts.* Daerah Istimewa Yogyakarta: Deepublish.

Nation, K. (2009). "Reading comprehension and vocabulary: What's the connection?" in *Beyond Decoding: The Behavioral and Biological Foundations of Reading Comprehension*, eds R. K. Wagner, C. Schatschneider, and C. Phythian-Sence (The Guilford Press), 176–194.

Nelson, H. E., and Willison, J. (1991). *National Adult Reading Test (NART).* Windsor: Nfer-Nelson, 1–26.

Nicholas, L. E., MacLennan, D. L., and Brookshire, R. H. (1986). Validity of multiple-sentence reading comprehension tests for aphasic adults. *J. Speech Hear. Disord.* 51, 82–87. doi: 10.1044/jshd.5101.82

OECD (2012). *Literacy, Numeracy and Problem Solving in Technology-Rich Environments: Framework for the OECD Survey of Adult Skills.* Paris: OECD Publishing.

OECD (2021). *OECD Skills Studies the Assessment Frameworks for Cycle 2 of the Programme for the International Assessment of Adult Competencies.* OECD Publishing.

OECD/Statistics Canada (1995). *Literacy, Economy and Society: Results of the First International Adult Literacy Survey.* Paris: OECD.

Olson, K., Smyth, J. D., Wang, Y., and Pearson, J. E. (2011). The self-assessed literacy index: Reliability and validity. *Soc. Sci. Res.* 40, 1465–1476. doi: 10.1016/j.ssresearch.2011.05.002

Ozuru, Y., Briner, S., Kurby, C. A., and McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Can. J. Experim. Psychol./Revue Canadienne de Psychologie Expérimentale* 67, 215–227. doi: 10.1037/a0032918

Pae, H. K., Greenberg, D., and Williams, R. S. (2012). An analysis of differential response patterns on the Peabody Picture Vocabulary Test-IIIB in struggling adult readers and third-grade children. *Read. Writ.* 25, 1239–1258. doi: 10.1007/s11145-011-9315-x

Patalay, P., Hayes, D., and Wolpert, M. (2018). Assessing the readability of the self-reported strengths and difficulties questionnaire. *BJPsych Open* 4, 55–57. doi: 10.1192/bjo.2017.13

Pellizzari, M., and Fichen, A. (2017). A new measure of skill mismatch: theory and evidence from PIAAC. *IZA J. Labor Econ.* 6. doi: 10.1186/s40172-016-0051-y

Perea, M., Rosa, E., and Gómez, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Mem. Cognit.* 30, 34–45. doi: 10.3758/BF03195263

Perry, K. H., Shaw, D. M., Ivanyuk, L., and Tham, Y. S. S. (2017). *Adult Functional Literacy: Prominent Themes, Glaring Omissions, and Future Directions*, 13.

Perry, K. H., Shaw, D. M., Ivanyuk, L., and Tham, Y. S. S. (2018). The 'ofcourseness' of functional literacy: ideologies in adult literacy. *J. Literacy Res.* 50, 74–96. doi: 10.1177/1086296X17753262

Pyrczak, F. (1972). Objective evaluation of the quality of multiple-choice test items designed to measure comprehension of reading passages. *Read. Res. Q.* 8, 62. doi: 10.2307/746981

Rain, M., and Mar, R. A. (2014). Measuring reading behavior: examining the predictive validity of print-exposure checklists. *Empir. Stud. Arts* 32, 93–108. doi: 10.2190/EM.32.1f

Reja, U., Manfreda, K. L., Hlebec, V., and Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Dev. Appl. Statist.* 19, 159–177.

Rose, S. J. (2009). "Identifying and teaching children and young people with dyslexia and literacy difficulties: An independent report from Sir Jim Rose to the Secretary of State for Children, Schools and Families," in *Department for Children, Schools and Families*.

Rouet, J. F., Britt, M. A., Gabrielsen, E., Kaakinen, J., Richter, T., and Lennon, M. (2021). *PIAAC Cycle 2 Assessment Framework: Literacy*.

Rupp, A. A., Ferne, T., and Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective. *Lang. Testing* 23, 441–474. doi: 10.1191/0265532206lt337oa

Sabatini, J., and Bruce, K. M. (2009). *PIAAC Reading Component: A Conceptual Framework*. OECD Education Working Papers, No. 33. EDU/WKP(2009)12. Paris: OECD Publishing.

Sabatini, J., O'Reilly, T., and Deane, P. (2013). Preliminary reading literacy assessment framework: foundation and rationale for assessment and system design. *ETS Res. Report Ser.* 2013, i–50. doi: 10.1002/j.2333-8504.2013.tb02337.x

Sabatini, J., O'Reilly, T., Dreier, K., and Wang, Z. (2019). "Cognitive processing challenges associated with low literacy in adults," in *The Wiley Handbook of Adult Literacy,* ed D. Perin (John Wiley & Sons), 15–39.

Schmidt, F. T., and Retelsdorf, J. (2016). A new measure of reading habit: Going beyond behavioral frequency. *Front. Psychol.* 7, 1364. doi: 10.3389/fpsyg.2016.01364

Schüller-Zwierlein, A., Mangen, A., Kovač, M., and van der Weel, A. (2022). Why higher-level reading is important. *First Monday* 27. doi: 10.5210/fm.v27i5.12770

Schutte, N. S., and Malouff, J. M. (2007). Dimensions of reading motivation: development of an adult reading motivation scale. *Read. Psychol.* 28, 469–489. doi: 10.1080/02702710701568991

Share, D. L. (2021). Is the science of reading just the science of reading English? *Read. Res. Q.* 56, S391–S402. doi: 10.1002/rrq.401

Shaywitz, S. E., and Shaywitz, B. A. (2005). Dyslexia (Specific Reading Disability). *Biol. Psychiatry* 57, 1301–1309. doi: 10.1016/j.biopsych.2005.01.043

Snowling, M., Dawes, P., Nash, H., and Hulme, C. (2012). Validity of a protocol for adult self-report of dyslexia and related difficulties. *Dyslexia* 18, 1–15. doi: 10.1002/dys.1432

Snyder, L., Caccamise, D., and Wise, B. (2005). The assessment of reading comprehension. *Top. Lang. Disord.* 25, 33–50. doi: 10.1097/00011363-200501000-00005

Stanovich, K. E., and West, R. F. (1989). Exposure to print and orthographic processing. *Read. Res. Q.* 24, 402. doi: 10.2307/747605

Stark, Z., Elalouf, K., Soldano, V., Franzen, L., and Johnson, A. (2023). *Validation and Reliability of the Dyslexia Adult Checklist in Screening for Dyslexia*. doi: 10.31234/osf.io/2r5ct

Stevens, K. B., and Price, J. R. (1999). Adult reading assessment: Are we doing the best with what we have? *Appl. Neuropsychol.* 6, 68–78. doi: 10.1207/s15324826an0602_2

Sticht, T. G. (2001). *The International Adult Literacy Survey: How Well Does it Represent the Literacy Abilities of Adults?*

Thompkins, A. C., and Binder, K. S. (2003). A comparison of the factors affecting reading performance of functionally illiterate adults and children matched by reading level. *Read. Res. Q.* 38, 236–258. doi: 10.1598/RRQ.38.2.4

Tighe, E. L., and Schatschneider, C. (2016). Examining the relationships of component reading skills to reading comprehension in struggling adult readers: a meta-analysis. *J. Learn. Disabil.* 49, 395–409. doi: 10.1177/0022219414555415

Tuinman, J. J. (1973). Determining the passage dependency of comprehension questions in 5 major tests. *Read. Res. Q.* 9, 206. doi: 10.2307/747135

Vágvölgyi, R., Bergström, K., Bulajic, A., Klatte, M., Fernandes, T., Grosche, M., et al. (2021). Functional illiteracy and developmental dyslexia: looking for common roots. *A systematic review. J. Cult. Cogn. Sci.* 5, 159–179. doi: 10.1007/s41809-021-00074-9

Vágvölgyi, R., Coldea, A., Dresler, T., Schrader, J., and Nuerk, H. C. (2016). A review about functional illiteracy: definition, cognitive, linguistic, and numerical aspects. *Front. Psychol.* 7, 1617. doi: 10.3389/fpsyg.2016.01617

West, R. F., Stanovich, K. E., and Mitchell, H. R. (1993). Reading in the real world and its correlates. *Read. Res. Q.* 28, 35–50. doi: 10.2307/747815

White, J., Burkhardt, A., Yeatman, J. D., and Goodman, N. (2022). "Automated generation of sentence reading fluency test items," in *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

Wicht, A., Durda, T., Krejcik, L., Artelt, C., Grotlüschen, A., Rammstedt, B., et al. (2021). Low "Literacy is not set in stone: longitudinal evidence on the development of low literacy during adulthood," in *Zeitschrift für Pädagogik,* 67. doi: 10.25656/01:28836

Wolff, U., and Lundberg, I. (2003). A technique for group screening of dyslexia among adults. *Ann. Dyslexia* 53, 324–339. doi: 10.1007/s11881-003-0015-3

Yagyu, K., Hashimoto, R., Shimojo, A., Iwata, M., Sueda, K., Seki, A., et al. (2021). Development of a reading difficulty questionnaire for adolescents in Japanese. *Brain Dev.* 43, 893–903. doi: 10.1016/j.braindev.2021.05.007

Yeatman, J. D., Tang, K. A., Donnelly, P. M., Yablonski, M., Ramamurthy, M., Karipidis, I. I., et al. (2021). Rapid online assessment of reading ability. *Sci. Rep.* 11, 6396. doi: 10.1038/s41598-021-85907-x

Zhang, S., Liu, X., Liu, J., Gao, J., Duh, K., and Van Durme, B. (2019). Record: bridging the gap between human and machine commonsense reading comprehension. *arXiv* preprint arXiv:1810.12885. doi: 10.48550/arXiv.1810.12885

Zhou, R., Wang, X., Zhang, L., and Guo, H. (2017). Who tends to answer open-ended questions in an e-service survey? The contribution of closed-ended answers. *Behav. Informat. Technol.* 36, 1274–1284. doi: 10.1080/0144929X.2017.1381165