Proficiency level descriptors for low reading proficiency: An integrative process model

Durda, Tabea; Artelt, Cordula; Lechner, Clemens; Rammstedt, Beatrice; Wicht, Alexandra

ORIGINAL PAPER

Proficiency level descriptors for low reading proficiency: An integrative process model

Tabea Durda[1] · Cordula Artelt[1] · Clemens M. Lechner[2] · Beatrice Rammstedt[2] · Alexandra Wicht[3]

Abstract

Low literacy among adolescents and adults worldwide remains a serious problem, even in economically developed countries. The consequences of low reading proficiency levels can be harmful in many ways for both the individuals concerned and their communities in terms of health, political, social and economic outcomes. While large-scale international assessments do assess reading proficiency, the data they provide for the bottom end of the scale are still somewhat undifferentiated. What is of particular concern to scholars and policymakers alike is to better understand the nature of reading difficulties among low-literate adolescents and adults. Addressing this need, the authors of this article present a new integrative *process model* which takes into consideration reader-related, text-related and task-related factors along different stages of the reading process that can cause reading difficulties. The process model incorporates different traditions of research on low reading proficiency: large-scale assessments, cognitive psychology, and research on developmental precursors of reading comprehension. It enabled the authors to identify core difficulty-generating factors, in particular task and text characteristics relevant in evaluating the difficulty of a reading task and thus in determining whether low-literate readers can solve it. The process model also proved suitable for incorporation into standard-setting practice. The authors demonstrate how the process model provided a framework for developing proficiency level descriptors (PLDs) which were then applied for the purpose of standard setting in a German large-scale assessment, the National Educational Panel Study (NEPS). Their results show that the PLDs were indeed suitable for differentiating between a *low* reading proficiency level and a *functional* reading proficiency level among adolescents and adults.

Keywords process model · low reading proficiency · large-scale assessments · difficulty-generating factors · standard setting

✉ Tabea Durda
  tabea.durda@lifbi.de

Extended author information available on the last page of the article

Résumé

Les descripteurs de niveaux appliqués à la mauvaise maîtrise de la lecture: une modélisation intégrative – Les faibles niveaux de littératie parmi les adolescents et les adultes du monde entier demeurent un grave problème, même dans les pays économiquement développés. Une mauvaise maîtrise de la lecture peut avoir des conséquences néfastes à maints égards, tant pour les personnes concernées que pour leurs communautés en termes de santé et de politique, et aux plans social et économique. Les évaluations réalisées à l'échelle internationale portent certes sur la maîtrise de la lecture, mais les données qu'elles fournissent au sujet des niveaux les plus faibles dans ce domaine sont encore quelque peu indifférenciées. Mieux comprendre la nature des difficultés qu'éprouvent les adolescents et adultes faiblement alphabétisés à lire recèle un intérêt particulier pour les universitaires et les décideurs politiques. En réponse à cette nécessité, les auteurs de l'article présentent une nouvelle *modélisation de processus* intégrative qui prend en compte des facteurs liés aux lecteurs, aux textes et aux tâches au fil de différentes étapes du processus de lecture, susceptibles de poser des difficultés. Cette modélisation englobe différentes approches scientifiques de la mauvaise maîtrise de la lecture : évaluations à vaste échelle, psychologie cognitive et recherche sur les précurseurs développementaux de la compréhension écrite. Elle permet aux auteurs d'identifier les facteurs essentiels occasionnant des difficultés, notamment les caractéristiques des textes et des exercices, utiles pour évaluer la difficulté d'un exercice de lecture et, ainsi, pour déterminer comment des lecteurs peu exercés peuvent les résoudre. Il s'avère aussi que la modélisation convient à une intégration dans une pratique normative. Les auteurs démontrent comment la modélisation a fourni un cadre permettant de développer des descripteurs de niveau, utilisés par la suite à des fins normatives dans le cadre d'une évaluation allemande à vaste échelle, l'étude longitudinale nationale sur l'éducation (National Educational Panel Study-NEPS). Les résultats ont montré que les descripteurs de niveaux convenaient effectivement pour faire la différence entre une *faible* maîtrise et un niveau de maîtrise *fonctionnelle* de la lecture chez les adolescents et les adultes.

Introduction

A significant number of adolescents and adults worldwide are only able to read at low proficiency levels, even in economically developed countries. The consequences of low reading proficiency levels can be harmful in many ways for both the individuals concerned and their communities in terms of health, political, social and economic outcomes (OECD 2012, 2013).

This issue is being addressed in the fourth of the United Nations Sustainable Development Goals (SDG 4), in particular in SDG target 4.6:

> By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy (UN 2016).

A profound understanding of the phenomenology and components of low reading proficiency levels across different large-scale assessments is relevant to the efforts

of achieving SDG target 4.6 in various ways. For example, in order to improve educational programmes and policies, policymakers should know to what extent adolescent students or adults identified as "low-literate" are comparable across studies. Moreover, it would be useful for practitioners such as teachers (not only those engaged in further education) to obtain guidelines for person-oriented interventions for formative assessment. Finally, test developers may need to construct or evaluate reading tasks based on theoretically justifiable criteria for low reading proficiency levels; for example for the purpose of describing and defining proficiency levels.

Although research on reading comprehension of adolescents and adults with low literacy proficiency levels has been conducted for decades, there is still no precise, consistent and generally accepted definition of what constitutes *low reading proficiency* (for reviews, see Eme 2011; Vágvölgyi et al. 2016). In public debate, the description of low reading proficiency is mainly derived from the description of the lowest proficiency level(s) in international comparative large-scale assessments such as the Programme for the International Assessment of Adult Competencies (PIAAC), where *low* refers to "Level 1" and "below Level 1" (OECD 2013),[1] the Programme for International Student Assessment (PISA), where *low* refers to "Level 1" (OECD 2012),[2] or the German Level-One Survey (LEO), where *low* refers to "Alpha Level 3" (Grotlüschen and Riekmann 2012).[3] All these assessments divide the continuous *latent* scale,[4] which measures a person's ability, into meaningful levels and describe each of them using *proficiency level descriptors* (PLDs), which represent specific requirements necessary to solve a task at a given proficiency level.

However, to obtain a precise understanding of low reading proficiency, further perspectives should be taken into account – for two reasons. First, different large-scale assessments implement distinct design parameters, for example regarding the population of interest (national, cross-national, adolescent, adult etc.) as well as in terms of literacy frameworks (focus on low proficiency or entire ability spectrum,

---

[1] PIAAC uses five proficiency levels, ranging from Level 1 to Level 5, with an additional "below Level 1" category at the bottom end of the scale. Individuals who score at Level 1 and below can read and locate a single piece of information in relatively simple texts in which the information requested is identical or synonymous with the information given in the question or directive. Adults at this proficiency level usually understand sentences or sections, but are generally unable to process, compare and evaluate several pieces of information. For descriptions of Levels 1–5, see OECD (2013, pp. 66–67).

[2] PISA 2012 uses seven proficiency levels, ranging from Levels 1b and 1a to Level 6. At the bottom end of this scale, individuals who score at Level 1b are able "to locate a single piece of explicitly stated information in a prominent position in a short, syntactically simple text with a familiar context and text type, such as a narrative or a simple list. The text typically provides support to the reader, such as repetition of information, pictures or familiar symbols. There is minimal competing information. In tasks requiring interpretation, the reader may need to make simple connections between adjacent pieces of information" (OECD 2012, p. 267).

[3] LEO, conducted in Germany to differentiate within the lowest proficiency levels, uses five so-called Alpha Levels based on a developmental approach to how written reading and writing skills are acquired. At the bottom end, the LEO scale begins with letters (Alpha Level 1); then progresses to words (Alpha Level 2); next to sentences (Alpha Level 3), next to short simple texts (Alpha Level 4) and finally to longer and more complex texts (Alpha Level 5/6) (Grotlüschen and Riekmann 2012)

[4] In this context, latent refers to a proficiency that cannot be observed directly, but is inferred from the directly observed variables (e.g. responses to reading tasks).

focus on writing and reading or reading only, paper-based or digital reading etc.) (e.g. Gehrer et al. 2013; Hartig and Riekmann 2012; Kirsch 2001; OECD 2012, 2013; Olsen and Nilsen 2017), which in turn might affect the description of low literacy in the respective large-scale assessment.

Second, large-scale assessments provide only limited information about the factors that underlie difficulties in solving a particular reading task. A better understanding of the cognitive processes and reading-related skills that lead to such difficulties would considerably enhance our comprehension of low reading proficiency in large-scale assessments. Although there is a great deal of research on this subject (e.g. Barth et al. 2015; Gernsbacher et al. 1990; Kintsch 1998; Long et al. 1999; Tighe and Schatschneider 2016), findings from these studies are largely unrelated to the perspective of large-scale assessments.

To close this gap, our article introduces a new *process model*[5] that explains low reading proficiency from an integrative perspective to obtain a comprehensive understanding of the reader-related, text-related and task-related factors along different stages of the reading process that can cause reading difficulties. Our article is structured into three parts:

In the first part, we outline the process model, including research on low reading proficiency from (1) large-scale assessments with a focus on task-text-reader characteristics; (2) cognitive psychology explaining proficiency differences according to underlying cognitive structures and processes relevant for text processing; and (3) research explaining proficiency differences according to developmental precursors of reading comprehension.

Second, based on this process model, we outline core difficulty-generating factors, in particular task and text characteristics that are relevant in evaluating the difficulty of a reading task and thus in determining whether low-literate readers can solve it.

In the third step, we illustrate the benefit of the process model for assessing low reading proficiency by incorporating the model into standard-setting practice. For this purpose, we outline how the process model provided the framework for developing PLDs used for standard setting in a German large-scale assessment, the National Educational Panel Study (NEPS, Blossfeld and Roßbach 2019), to differentiate between low-literate and functionally literate adolescents and adults.[6]

---

[5] The details of this process model are discussed later in this article. For now, in a nutshell, a process model typically describes the activities or steps that contribute to accomplishing a specific goal or task. In our case, the process model describes the steps a reader takes within a reading process and the factors that may facilitate or hinder the reading process.

[6] The rationale for this part also emerges from our own involvement in a project entitled "Identifying risk and protective factors for the development of low literacy and numeracy among German adults" (*Risiko- und Schutzfaktoren für geringe Literalität und Numeralität bei deutschen Erwachsenen;* GeLiNu), for which a threshold or "cut score" for low literacy had to be derived. For more information about this project (2017–2020), visit https://www.gesis.org/en/research/external-funding-projects/overview-external-funding-projects/gelinu [accessed 30 March 2020].

A process model for explaining difficulties in the accomplishment of reading tasks among low-literate readers

Reading comprehension involves more than the ability to read. It includes a variety of reading activities that rely on several developmental precursors of reading comprehension (e.g. word recognition, comprehension, working memory) as well as on cognitive and metacognitive[7] strategies that together help a reader to understand, process, restructure, evaluate and monitor information to accomplish all kinds of reading requirements (e.g. Artelt et al. 2005; Eme 2011; Tighe and Schatschneider 2016). In task-oriented reading situations, such as in test situations of large-scale assessments, those activities differ depending on the stage of the reading process, the specific requirements of the task and the characteristics of the written material readers are confronted with (Cerdán et al. 2011; Rouet 2006).

Roughly speaking, the reading process comprises three stages: In the *preparation stage,* the reader builds a task model in which he or she determines the purpose of the task. In the subsequent *execution stage,* the reader needs to find relevant information in the text and, if necessary, he or she must relate several pieces of information to each other. Within this process, the reader has to evaluate the relevance of each piece of information and then decide whether to maintain or reject it. This verification and integration process may require several cycles. The reader may also have to revise his or her task model. Finally, in the *production stage,* the reader needs to verify once again whether the selected information suffices to answer the question and if so, he or she provides a response (see Figure 1).
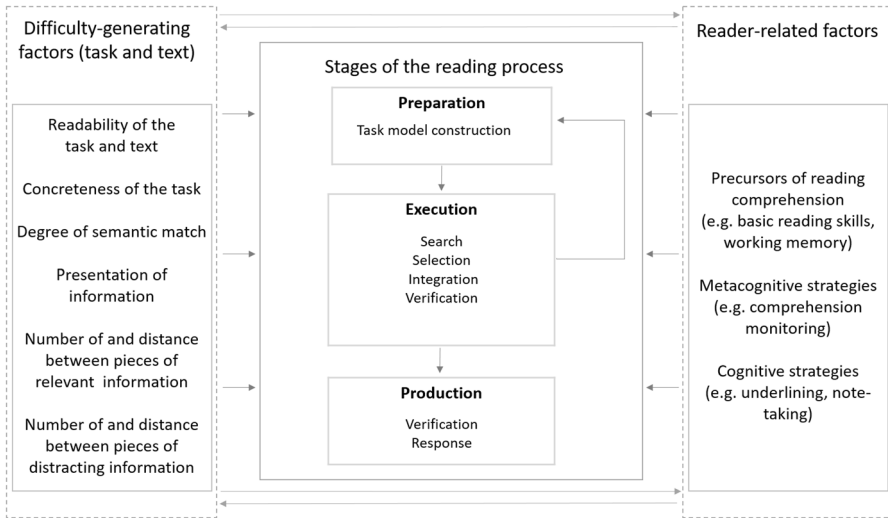
Difficulties in the accomplishment of reading tasks among low-literate readers can be traced along these stages of the task-oriented reading process (Kirsch and Mosenthal 1990; Rouet 2006). Therefore, in the next three sections, we sketch the challenges that low-literate readers might face during (1) the construction of a task model, (2) the process of searching for and selecting information and (3) the process of integrating information along with task-related, text-related and readers' characteristics.

Difficulties among low-literate readers in task model construction

Reading tasks in large-scale assessments often aim to reflect realistic reading requirements that occur in a variety of reading contexts. Such tasks might require a reader to locate a single piece of information or to draw simple inferences. Other tasks might tap requirements like locating several pieces of information, integrating widely distributed information and reflecting on the intention, content or form of the text (e.g. Gehrer et al. 2013; Kirsch 2001; OECD 2012, 2013). In this context, research has identified several difficulty-generating factors that relate to the wording of the reading task (question and given options), the task format or the reading

---

[7] Cognition refers to the mental processes involved in acquiring knowledge and comprehension. Metacognition, or "thinking about thinking", refers to the knowledge and awareness of one's own cognition and the regulation of cognition.

Fig. 1 Proposed process model for explaining difficulties in the accomplishment of reading tasks among low-literate readers

requirements. In this article, we focus on the wording and the reading requirements because of their particular importance for people with low reading proficiency.

In terms of wording, to understand what should be done in the execution stage (the second stage of the reading process), a reader must first extract the given and requested information from the reading task. This requires the coordination of multiple basic reading skills, ranging from word retrieval and syntactic parsing[8] to comprehension of the meaning of the information (Perfetti and Hart 2002; Perfetti and Stafura 2014). Low-literate readers are often portrayed as having lower basic reading skills than more proficient readers (e.g. Eme 2011; Landi 2010; McKoon and Ratcliff 2018; Tighe and Schatschneider 2016). Therefore, differences in goal formation (task model construction) can be traced to the wording of the task. In particular, it can be assumed that low-literate readers are more likely to struggle with understanding the reading requirements when tasks contain more *low-frequency words* and are *propositionally denser*[9] (e.g. Embretson and Wetzel 1987; Hartig and Frey 2012; Hartig and Riekmann 2012; Kintsch and van Dijk 1978; Ozuru et al. 2008; Sonnenleitner 2008; Zimmermann 2016).

---

[8] Word retrieval refers to the reader's activity of finding (retrieving) words from their lexicon that they have stored in their memory. Syntactic parsing refers to the reader's activity to construct a representation of the syntactic structure of each sentence.

[9] Low-frequency words are words which do not occur very often in everyday spoken or written language. (e.g. *purchase, albeit*). The opposite are high-frequency (commonly used) words (e.g. *the, yes, no*). Propositional density refers to the number of idea units, or propositions relative to the total number of words. The greater the density, the larger the cognitive load on the reader.

Regarding reading requirements, existing studies stress that each type of reading question evokes different reading activities that readers will execute to solve a reading task (Cerdán et al. 2019; Rouet et al. 2001). According to Jean-François Rouet and colleagues (2001, p. 175), reading questions that require a reader to locate single pieces of information promote "locate-and-memorize" reading strategies, whereas questions that require integrating information or building up a *situation model*[10] promote "review-and-integrate" reading strategies. Therefore, another potential explanation for proficiency differences is that low-literate readers are less likely to construct an appropriate task model than proficient readers. This may impede them setting themselves appropriate reading goals regarding which information to search for and what type of answer to provide in order to progress through the text and complete the task.

This assumption is supported by a number of experimental studies which demonstrated that readers with lower reading abilities are less likely than proficient readers to construct an appropriate task model that helped them to adapt their reading process (e.g. Cerdán et al. 2011, 2019; de Milliano et al. 2016). For example, Raquel Cerdán and colleagues (2019) designed an experiment in which they manipulated the reading question given to participants. Their aim was to test whether low-literate and literate readers differed in their reading comprehension depending on whether or not the reading question was paraphrased to highlight the required cognitive processes to answer the question. For example, the reading question

Miguel works in the ACOL Company and in the 17th of May week will be on a business trip. Justify if Miguel should contact Raquel.

was paraphrased into

Explain if Miguel should get in touch with Raquel to get the vaccine and why (Cerdán et al. 2019, p. 2117).

Cerdán and colleagues (2019) found that only low-literate readers, but not proficient readers, benefited from such cues. Similar results were found with respect to the wording of test instructions for students with special educational needs in Germany (e.g. Nusser and Weinert 2017). This indicates both a poorer situation model of the reading task among low-literate readers and the dependence of a task's difficulty on the *concreteness of the question* concerning the required cognitive processes.

Difficulties among low-literate readers in searching for and selecting information

Once readers have constructed a task model, they have to search for, select and process relevant parts of the text according to the task requirements, irrespective of whether only one piece of information needs to be located or several pieces of information need to be integrated (Cerdán et al. 2011; Rouet 2006). How easily and

---

[10] A situation model refers to a mental representation of the text; it includes prior knowledge in addition to the information from the text. With the construction of such a mental model, the reader has not only understood what is said, but can also "imagine" something.

successfully this can be accomplished depends heavily on the extent of a reader's cognitive effort to select the relevant information. The cognitive effort, in turn, seems to depend particularly on *semantic cues* at the micro level and *text-signalling devices* at the macro level[11] that aid a reader in drawing his/her attention to where the relevant information can be found (Kintsch 1998; White 2012).

At the micro level, it is well established that the cognitive complexity of searching for and selecting relevant information is influenced by the type of semantic match between the information in the text and in the reading task (e.g. Kintsch 1998; Lumley et al. 2012; OECD 2012, 2013; Todaro et al. 2010; White 2012; Zimmermann 2016). In its simplest form, the match is facilitated by a literal overlap that activates the same semantic unit, called a *semantic proposition* (Kintsch 1998). In practical terms, this means, for example, that reading tasks where key words appear in both the reading question/or options and in the headline of the text are easier than tasks where a reader has to infer the relationship. To give an example (Yang et al. 2005, p. 235):

> After being dropped from the plane, the bomb hit the ground and exploded.
> The explosion was quickly reported to the commander [literal match].

compared to

> After being dropped from the plane, the bomb hit the ground. The explosion
> was quickly reported to the commander [inference].

It has been shown that reading tasks in large-scale assessments at the lowest proficiency levels require a reader to retrieve information from the text which is for the most part literally identical to the information in the reading task more frequently than reading tasks for higher proficiency levels. By contrast, tasks at the second-lowest proficiency level more often require paraphrasing or inferring (Kirsch 2001; OECD 2012, 2013).

One possible explanation is that low-literate readers require more cognitive resources to comprehend the meaning of the information in the text because the reading processes at their disposal are less automated in terms of word retrieval and syntactic parsing (McKoon and Ratcliff 2018; Yang et al. 2005). This lack of automated processes, in turn, leaves less processing capacity for other types of matching, such as drawing inferences (Perfetti and Stafura 2014). However, studies have also demonstrated that a literal match can be misleading and can increase difficulty, namely when not the relevant but rather the irrelevant information (also termed *distracting information*) matches with that in the task. For example, Cerdán and colleagues (2011) showed in an experiment that low-literate readers tended to select information in the text that matched information in the task (question):

---

[11] Semantic refers to meaning in language. Semantic cues provide hints to facilitate inferring word meanings (e.g. a semantic cue for the word "knife" might be "It's used to cut food"). Text-signalling devices make information stand out in print. Examples are a headline; a blank line before important information; italics, bold print or coloured or larger font; or symbols like a bullet point or an arrow.

Which adverse reactions can the vaccine provoke? Those students who would be falsely seduced by superficial cues would tend to visit the following distracting location of the text: Ask your doctor if you are under medical treatment or have had adverse reactions to the Flu vaccine. If these are very intense it may be dangerous to the Fetus in case of pregnancy (Cerdán et al. 2011, pp. 203–204).

Because of this strategy, low-literate students were more often misled by distracting information. Corresponding results from large-scale assessments have shown the difficulty-generating effect of the amount of distracting information in the text (Kirsch 2001; Lumley et al. 2012; Ozuru et al. 2008; Zimmermann 2016). The tasks at the lowest proficiency levels hardly contained any distracting information, whereas the amount of distracting information increased at the second-lowest proficiency level (Hartig and Frey 2012; Kirsch 2001; OECD 2012, 2013).

Low-literate readers further differ from more proficient readers in terms of the efficiency with which they use adequate reading strategies that help them in searching for and selecting task-relevant information (Cataldo and Oakhill 2000; Cerdán et al. 2009; Hahnel et al. 2018). For example, Giulia Maria Cataldo and Jane Oakhill (2011) showed that low-literate secondary students more often used an inefficient and undirected reading strategy, such as rereading the entire text when they were required to locate specific pieces of information, compared to skilled readers who were more capable of locating the relevant passage. Therefore, the second explanation in proficiency differences refers to the *presence or absence of cues at the macro level with respect to text-signalling devices*. Studies have shown that low-literate readers use headlines or bullet points rather than larger, more sustained text segments to retrieve information (DelVecchio et al. 2019; Weeks 2001). Moreover, research in the field of health communication has demonstrated that texts with additional visual information (Braich et al. 2011; Dowse and Ehlers 2005; van Beusekom et al. 2016) or with typographical devices, such as coloured font, capital letters or underlining (Bass et al. 2016), helped low-literate readers to find and recall information better than textual information alone. Corresponding results from large-scale assessments also found that low-literate readers were more capable of locating and selecting relevant information when it was prominently placed, such as in the heading or at the beginning of the text, or stood out due to other features such as the presence of a number in the text (OECD 2012, 2013).

Difficulties among low-literate readers with integrating information

In many reading tasks, readers must often integrate several pieces of information (e.g. for comparing statements or understanding the author's intention). There is ample research showing that low-literate readers struggle with integrating information (e.g. Barth et al. 2015; Gernsbacher et al. 1990; Kirsch 2001; Long and Chong 2001; Magliano and Millis 2003; OECD 2012, 2013). Some explanations have already been provided (see the two earlier sections, Difficulties among low-literate readers in task model construction and Difficulties among low-literate readers in searching for and selecting information).

Another explanation relates to the *number of* and *distance between pieces of information relevant* to the reading task. Results from large-scale assessments indicate that whereas more proficient readers are more likely to integrate two or more pieces of information distributed across the text, low-literate readers are more likely to locate single pieces of information and integrate more than one piece of information located in neighbouring sentences (OECD 2012, 2013; Kirsch 2001). Similarly, studies from cognitive psychology suggest that low-literate readers have greater difficulty in drawing inferences using distributed pieces of information (*global inference*) than from information found in neighbouring sentences (*local inference)* (Barth et al. 2015; Long and Chong 2001; Gernsbacher et al. 1990; Magliano and Millis 2003). This seems to be especially due to differences in working memory capacities (e.g. Abadzi 2008; Carretti et al. 2009). Studies have shown that low-literate readers can only keep a small number of pieces of information active in their short-term memory and reactivate it when they read; both of these activities are important prerequisites for integrating pieces of information (Carretti et al. 2009; Chiappe et al. 2000; De Beni et al. 1998; Mellard and Fall 2012). In particular, low-literate adults might struggle to integrate information provided across larger text segments because the relevant corresponding information is no longer available in their short-term memory but must be retrieved from their long-term memory. This is coupled with the finding that they have more difficulties in suppressing irrelevant information (Carretti et al. 2009; Daneman and Carpenter 1980; Gernsbacher et al. 1990; Long et al. 1999).

In addition, difficulties can be traced to readers' metacognitive and cognitive strategies. One important metacognitive skill is to monitor the comprehension process itself, since the detection of inconsistencies helps a reader to adapt his or her reading behaviour to restore coherence (Helder et al. 2016). Thus, a successful integration of information depends on how well the reader succeeds in monitoring the coherence of new pieces of information with those that are already activated. In this context, studies have shown that low-literate readers are less sensitive to detecting coherence breaks than literate readers (Barth et al. 2015; Long and Chong 2001; Todaro et al. 2010). Investigations into how much the distance between relevant pieces of information mattered found that low-literate readers were more likely to detect local coherence breaks, but less likely to detect inconsistencies when contradictory pieces of information were separated by intervening sentences (Barth et al. 2015; Long and Chong 2001). Regarding cognitive reading strategies, studies emphasise that, unlike proficient readers, low-literate readers use fewer efficient reading strategies that would allow them to retain pieces of information more easily and then relate them to each other. For example, compared to more proficient readers, low-literate readers less often used strategies such as underlining, note-taking and summarising key information that would have helped them to capture, retrieve and restructure information (Artelt et al. 2001; Cromley and Azevedo 2007; de Milliano et al. 2016).

Establishing difficulty-generating factors

In this section, we demonstrate how the process model outlined earlier in this article makes it possible to identify the main difficulty-generating factors that underlie proficiency differences between low-literate and literate readers. The process model (Figure 1) highlights that the difficulty of solving a reading task depends on the extent to which the task and stimulus text demand different reading-related skills.

First, whether or not low-literate readers adequately progress through the text depends on the *readability of the reading task* (e.g. in the form of high-frequency words) and the *concreteness of the wording* in terms of how clearly the processing steps required to answer the reading question are communicated (e.g. whether the question highlights or merely vaguely formulates the key information to be sought).

Second, the model emphasises that the success of low-literate readers in locating the relevant information depends on both the degree to which the required information matches the information given in the text, and the degree of clarity with which the text draws attention to the relevant passage in which the information can be found. However, in the absence of such cues, low-literate readers need to apply more cognitive resources and strategies to locate the information and to draw inferences about its relevance. Therefore, both the *degree of semantic match* and the *presentation of information* are further difficulty-generating factors.

Third, the *number of* and *distance between relevant pieces of information* affect whether or not low-literate readers can solve the reading task. The more pieces of information there are, and the larger the gap between the relevant pieces of information, the more demanding the task becomes in terms of basic reading (e.g. larger amount of syntactic parsing), comprehension (e.g. understanding and evaluating whether the piece of information is relevant or not) and processing (e.g. using appropriate reading strategies to integrate multiple pieces of information, thus suppressing irrelevant information).

Fourth, the *number of* and *distance between distracting pieces of information* can cause reading difficulties. Low-literate readers are more likely to erroneously choose answers that overlap with the distracting information. This seems to be particularly problematic when the distracting information appears before or near the relevant information.

Fifth, not only the readability of the reading task (the question and options) but also the *readability of the text* can explain why low-literate readers have difficulties in solving the task. For example, if the task requires a reader to locate a single piece of information and the text provides neither semantic cues nor text-signalling devices, then it is likely to be much more demanding to progress through a more complex text (e.g. in terms of propositional density) than through other texts with a lower complexity.

Standard setting in the German National Educational Panel Study (NEPS)

In this section, we describe how this process model and the identified difficulty-generating factors were applied for the purpose of standard setting in a German large-scale assessment, the National Educational Panel Study (NEPS; Blossfeld and Roßbach 2019). The aim was to establish a proficiency level for differentiating between low-literate and functionally literate adolescents and adults. We begin with a brief overview of NEPS and the assessment of reading proficiency before presenting the steps of the standard-setting procedure.

Overview of NEPS

NEPS provides rich longitudinal data on the development of competencies (reading, mathematics, scientific literacy, and information and communication literacy) following a representative *multi-cohort sequence design*[12] with six cohorts ranging from early childhood to adulthood. Thus, it provides a rich data source that enables scholars and practitioners to understand how learning environments shape competence development (e.g. among low-literate readers) and how competencies are related to educational decisions and returns to education in formal, non-formal and informal contexts throughout the lifespan (Blossfeld and Roßbach 2019).[13]

Briefly, the six NEPS *starting cohorts* (SCs)[14] are

SC 1   Starting Cohort Newborns
SC 2   Starting Cohort Kindergarten
SC 3   Starting Cohort Grade 5
SC 4   Starting Cohort Grade 9
SC 5   Starting Cohort First-Year University Students
SC 6   Starting Cohort Adults

In this article, we focus on two starting cohorts, Grade 9 adolescents (SC 4) and adults (SC 6), to apply a general understanding of the usability of the process model.[15] The ninth graders (SC 4, n = 13,897) attended regular schools (lower

---

[12] In terms of saving time, it is more efficient to have several cohorts (multi-cohort design) rather than only one. In addition, "concentrat[ing] on important sequences in the educational career" (Blossfeld et al. 2019, p. 11), i.e. having a sequence design, also enables to have comparable research results at an early stage in a longitudinal study.

[13] NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in Germany. For more information, visit the dedicated website at https://www.neps-data.de/Mainpage [accessed 31 March 2020].

[14] In a longitudinal study such as NEPS, a starting cohort is a cohort sample selected at the beginning of the study. The cohorts are either age-based (e.g. SC 6) or defined by specific stages in their educational career (e.g. SC 5).

[15] The dedicated online landing pages for these two cohorts are https://doi.org/10.5157/neps:sc4:10.0.0 and https://doi.org/10.5157/neps:sc6:10.0.0.

secondary school, intermediate secondary school, upper academic school track, comprehensive school and multi-track schools) in Germany and were, on average, 14.74 years old ($SD = 0.73$). Among the SC 4 sample, 49.14 per cent are female and 12.37 per cent have a non-German background. The students were first tested for their reading competence in 2011/12. Students with special educational needs are not included in the sample. The age of the adult sample (SC 6, $n = 8{,}480$) ranged from 24 to 69 years. They were, on average, 47.82 years old ($SD = 11.68$). Among the SC 6 sample, 48.62 per cent are female and 12.68 per cent have a non-German background. The participants of SC 6 were first tested for their reading competence in wave 3, 2010/11, or wave 5, 2012/13, by applying identical assessments (Haberkorn et al. 2012; Koller et al. 2014).

NEPS reading tests

Being a longitudinal study, NEPS aims to track the development in reading comprehension among other competencies over the lifespan. Therefore, the reading tests for all cohorts are based on the same framework, but they are adapted for age-appropriate texts and text topics (Gehrer et al. 2013). The NEPS reading framework considers three main dimensions, namely text functions, cognitive requirements and task formats. Each participant in the two cohorts under consideration here received five texts, each representing a different *text function*, including only continuous texts:[16] (1) an information text; (2) a commenting text or an argumentative text; (3) a literary text; (4) an instructional text; and (5) an advertising text. *Cognitive requirements* refer to the process participants must engage in to solve the task and are classified into finding information in the text, drawing text-related conclusions and reflecting and assessing (situation model). In terms of *task formats,* most of the items are presented in a multiple-choice format. Further task formats are decision-making items and matching items. Participants were provided with a standardised instruction and sample items to ensure that they understood the item formats (Gehrer et al. 2013; Haberkorn et al. 2012; Hardt et al. 2013; Koller et al. 2014).

The NEPS reading tests which we used for differentiating between a low reading proficiency level and a functional reading proficiency level consist of 31 items (SC 4) and 30 items (SC 6) respectively. To estimate reading competence, the NEPS data were scaled on *item response theory* (IRT),[17] showing a good reliability in SC 4 (WLE reliability = 0.749)[18] (Haberkorn et al. 2012) and in SC 6 (WLE reliability = 0.717/0.743 in wave 3/5) (Hardt et al. 2013; Koller et al. 2014). *Differential item functioning* (DIF) was tested for several population indicators (e.g., gender, school

---

[16] Continuous texts are also known as linear texts in contrast to non-continuous texts that are organised differently (e.g. tables, lists, diagrams).

[17] Item response theory refers to a family of measurement models that attempt to explain the relationship between a latent ability (i.e. unobservable ability, for example reading proficiency) and their manifestations (i.e. observed responses, e.g. responses to reading items).

[18] NEPS provides Warm's Mean Weighted Likelihood Estimates (WLE; Warm 1989) of a persons' reading ability given their item responses.

certificate, migration background) to evaluate the test fairness across several sub-groups. The analyses showed that the tests were fair for the examined subgroups, indicating that differences in reading competences and thus the probability of being assigned to the low-literate group are not due to differences in the item difficulties between subgroups such as native and non-native speakers (Haberkorn et al. 2012; Hardt et al. 2013; Koller et al. 2014). Participants were tested by trained interviewers either at their schools (SC 4) or in their homes (SC 6) via a paper-and-pencil test with a maximum test duration of 28 minutes.

Determining a threshold for low reading proficiency

Reports on (cross-sectional) international large-scale assessments regularly offer "cut-offs" (thresholds) for proficiency levels across the entire ability spectrum, including a more or less theoretically motivated definition of low reading proficiency at the lower levels (e.g. Grotlüschen and Riekmann 2012; OECD 2012, 2013; Olsen and Nilsen 2017). Because of the multi-cohort sequence design implemented in NEPS, however, the definition of proficiency levels for repeated assessments within and across cohorts is far more ambitious, particularly with respect to linkage across time and cohorts.

All parties involved in the project agreed that the standard-setting procedure for the definition of low reading proficiency within NEPS needed to follow an *a priori* theory-driven approach (from the outset); first because of the multi-cohort sequence design, and second because the NEPS reading literacy framework has conceptual overlaps with those of other large-scale assessments such as PIAAC and PISA (OECD 2012, 2013), but also differences. Overlaps include, for example, considering the entire reading ability spectrum, cognitive requirements and function of the texts. Conceptual and methodological differences include, for example, the breadth of texts such as non-continuous texts and digital texts, and differences in terms of scaling procedures.

The project team decided that the *Bookmark* method (Mitzel et al. 2012), due to its logical appeal and practicality (Karantonis and Sireci 2006), would be best suited to determining a threshold. It is one of the most frequently used standard-setting methods, and its approach involves subjective judgments to determine proficiency levels. In this method, the items are presented according to their item difficulty in an *ordered item booklet* (OIB), beginning with the easiest item. It is then the task of the panellists to place a cut-score "bookmark" between those items which, in their view, define the boundary between two proficiency levels (e.g. Cizek et al. 2005; Karantonis and Sireci 2006; Mitzel et al. 2012). Once the method had been chosen, the following two steps were carried out:
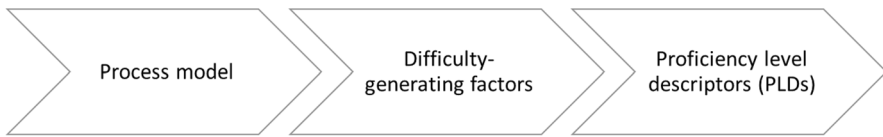
Fig. 2 Steps for developing proficiency level descriptors (PLDs)

Step 1: Translation of the difficulty-generating factors into proficiency level descriptors (PLDs)

Based on the process model, we translated the identified difficulty-generating factors into PLDs that served as guidance for the Bookmark Method (see Figure 2).

As shown in Table 1, the PLDs differentiate among the dimensions of the difficulty-generating factors. Since we expected the reading abilities of participants who were classified as literate readers to cover a wide range of achievements, we decided not to compare what individuals at the lowest proficiency level and those at the next highest proficiency level could do, but to contrast instead what low-literate readers can probably do and not do.

Step 2: Determining a proficiency level with the Bookmark method

First, we ordered the reading items in an OIB according to their item difficulty in the IRT models (Haberkorn et al. 2012; Hardt et al. 2013; Koller et al. 2014), beginning with the easiest item. Next, we briefed the panellists on how to set the cut score by explaining the Bookmark method and discussing the PLDs. The group of panellists consisted of experts, including test developers and professionals, who were working with large-scale assessments and reading comprehension tests in Germany. After this training, the Bookmark method was applied in three rounds (Cizek et al. 2005; Mitzel et al. 2012). Within those rounds, the panellists, in repeated comparison of the reading items with the PLDs, had to set a cut score "bookmark" between those items which, in their view, defined the boundary between a low reading proficiency level and a literate reading proficiency level. In the third round, the final cut score was determined. Based on this standard-setting procedure, 4.24 per cent of adolescents (SC 4, weighted) and 14.25 per cent of adults (SC 6, weighted) were assigned to the low-literacy group (for further results, see Wicht et al. forthcoming).

Discussion

This article presents an integrative process model we developed drawing on different research traditions – large-scale assessments, research on underlying cognitive processes and developmental precursors of reading (e.g. Barth et al. 2015; Gernsbacher et al. 1990; Hartig and Riekmann 2012; Kintsch 1998; Kirsch 2001; Long

Table 1 Proficiency level descriptors (PLD) of low-literate adolescents and adults within NEPS

Difficulty-generating factors	What low-literate readers *can* probably do	What low-literate readers probably *cannot* do
Readability	• solve a task which does not require detailed comprehension of the task or text, with only a basic understanding of the same being sufficient	• solve a task which requires the reader to demonstrate an accurate understanding of the task, or a text whose content is more complex (e.g. featuring low-frequency words, higher propositional density)
Degree of semantic match	• retrieve required information which is, for the most part, literally identical with the information in the task	• solve a task in which the required information needs to be inferred
Presentation of information	• use text-signalling devices (e.g. because of the presence of head-lines or bullet points) to find the information needed	• use text-signalling devices (e.g. because of the absence of headlines or bullet points) to find the information needed
Number of and distance between pieces of relevant information	• locate single pieces of information • cycle through more than one piece of information between neigh-bouring sentences (drawing local inferences)	• integrate two or more pieces of information • cycle through more than one piece of information, with information spread across the text (drawing global inferences)
Number of and distance between pieces of dis-tracting information	• accomplish tasks with minimal distracting information, if this minimal distracting information in the stimulus text is presented after the solution	• accomplish tasks with more distracting information, where this distracting information may be located in a previous text section, in the same section as the solution or in the following text section

et al. 1999; OECD 2012, 2013). Our purpose in developing it was to obtain a comprehensive understanding of the reader-related, text-related and task-related factors that explain reading difficulties along different stages of the reading process. Hence, our process model bridges the gap between factors proven to be related to low reading proficiency in large-scale assessments and factors proven to explain reading proficiency differences based on underlying processes and developmental precursors of reading comprehension. The combination of these perspectives offers a comprehensive explanatory model for low reading proficiency that can be applied to different educational contexts.

For example, as we demonstrated in the translation of the process model into difficulty-generating factors and subsequent application of the results to the standard-setting process within NEPS (Blossfeld and Roßbach 2019), the advantage of the chosen *a priori* approach is that the description of the proficiency levels has a stronger theoretical basis than with a purely *post hoc* (retrospective) interpretation of the item characteristics used in many comparative large-scale assessments (OECD 2012, 2013; Olsen and Nilsen 2017). The connection between the *a priori* developed PLDs and the item difficulties, determined independently of each other, can be regarded as a *construct validation* (Egan et al. 2012).[19] Furthermore, the difficulty-generating factors we identified using the process model are useful for the comparison of large-scale assessments. For example, it emerges that there is a high similarity in the descriptions of the low reading proficiency levels of NEPS, PISA (OECD 2012), PIAAC (OECD 2013) and in parts of the Progress in International Reading Literacy Study (PIRLS)[20] (Bos et al. 2012). However, further research is required to validate comparability.

At the same time, the process model also provides a template to question how valid and reliable the assignment of test participants to the low-literacy group is. For example, we have argued that the readers' skills, the task requirements and factors related to the text are relevant to understand reading difficulties among low-literate readers. To follow up on this interdependence, it is worth investigating the extent to which cognitive and language-related deficits explain assignment to the low-literacy group, especially if the readers are non-native speakers, in order to validate that a person is assigned to the low-literacy group because of difficulties in reading only (e.g. Abadzi 2008; Vágvölgyi et al. 2016).

Moreover, our process model provides some hints regarding a process-oriented assessment of a person's reading difficulties that can guide teaching measures. Through the model, the assessment process could start with an examination of a person's reading profile to assess whether the weaknesses lie, for example, primarily in the basic reading skills or rather in the processing skills (e.g. working memory or strategic skills). After having a clear picture of the individual's needs, the teacher could then assess which text-related and task-related factors might cause additional difficulties; for example, he or she might probe whether the difficulties result from

---

[19] Construct validation refers to the degree to which a test measures what it claims to measure.

[20] PIRLS, run by the International Association for the Evaluation of Educational Achievement (IEA) since 2001, assesses Grade 4 students' reading comprehension both in and out of school. For more information, visit https://www.iea.nl/studies/iea/pirls [accessed 2 April 2020].

the degree of concreteness of the reading task or the text structure. To ensure a comprehensive assessment, it might be helpful to probe the reading difficulties with different task formats, text forms (e.g. in terms of text types or text-signalling devices) and task requirements.

References

Abadzi, H. (2008). Efficient learning for the poor: New insights into literacy acquisition for children. *International Review of Education, 54*(5–6), 581–604. https://doi.org/10.1007/s11159-008-9102-3.

Artelt, C., McElvany, N., Christmann, U., Richter, T., Groeben, N., Köster, J., Schneider, W., Stanat, P., Ostermeier, C., Schiefele, U., Valtin, R., & Ring, K. (2005). *Expertise: Förderung von Lesekompetenz* [Expertise: Promotion of reading skills]. Bildungsreform Bd. 17. Berlin: Bundesministerium für Bildung und Forschung (BMBF). Retrieved 24 March 2020 from https://www.bildungsserver.de/onlineressource.html?onlineressourcen_id=31410.

Artelt, C., Schiefele, U., & Schneider, W. (2001). Predictors of reading literacy. *European Journal of Psychology of Education, 16*(3), 363–383. https://doi.org/10.1007/BF03173188.

Barth, A. E., Barnes, M., Francis, D., Vaughn, S., & York, M. (2015). Inferential processing among adequate and struggling adolescent comprehenders and relations to reading comprehension. *Reading and Writing, 28*(5), 587–609. https://doi.org/10.1007/s11145-014-9540-1.

Bass, S.B., Gordon, T.F., Gordon, R., & Parvanta, C. (2016). Using eye tracking and gaze pattern analysis to test a "dirty bomb" decision aid in a pilot RCT in urban adults with limited literacy. *BMC Medical Informatics and Decision Making, 16*, Art. 67. https://doi.org/10.1186/s12911-016-0304-5

Blossfeld, H.-P., & Roßbach, H.-G. (Eds). (2019). *Education as a lifelong process: The German National Educational Panel Study (NEPS).* Edition ZfE, vol. 3. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-23162-0

Blossfeld, H.-P., von Maurice, J., & Scheider, T. (2019). The National Educational Panel Study: Need, main features, and research potential. In H.-P. Blossfeld & H.-G. Roßbach (Eds), *Education as a lifelong process. The German National Educational Panel Study (NEPS).* Edition ZfE, vol 3 (pp. 1–16). Wiesbaden: Springer VS, https://doi.org/10.1007/978-3-658-23162-0_1

Bos, W., Taelli, I., Bremerisch-Vos, & Schwippert, K. (Eds.). (2012). *IGLU 2001: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich [Reading skills of primary school children in Germany international comparison].* Münster: Waxmann.

Braich, P. S., Almeida, D. R., Hollands, S., & Coleman, M. T. (2011). Effects of pictograms in educating 3 distinct low-literacy populations on the use of postoperative cataract medication. *Canadian Journal of Ophthalmology, 46*(3), 276–281. https://doi.org/10.1016/j.jcjo.2011.05.004.

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences, 19*(2), 246–251. https://doi.org/10.1016/j.lindif.2008.10.002.

Cataldo, G. M., & Oakhill, J. (2000). Why are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on the ability to locate information in text. *Journal of Educational Psychology, 92*(4), 791–799. https://doi.org/10.1037/0022-0663.92.4.791.

Cerdán, R., Gilabert, R., & Vidal-Abarca, E. (2011). Selecting information to answer questions: Strategic individual differences when searching texts. *Learning and Individual Differences, 21*(2), 201–205. https://doi.org/10.1016/j.lindif.2010.11.007.

Cerdán, R., Pérez, A., Vidal-Abarca, E., & Rouet, J. F. (2019). To answer questions from text, one has to understand what the question is asking: Differential effects of question aids as a function of comprehension skill. *Reading and Writing, 32*(8), 2111–2124. https://doi.org/10.1007/s11145-019-09943-w.

Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction, 19*(1), 13–27. https://doi.org/10.1016/j.learninstruc.2007.12.003.

Chiappe, P., Siegel, L. S., & Hasher, L. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition, 28*(1), 8–17. https://doi.org/10.3758/BF03211570.

Cizek, G. J., Bunch, M. B., & Koons, H. (2005). Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice, 23*(4), 31–50. https://doi.org/10.1111/j.1745-3992.2004.tb00166.x.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*(2), 311–325. https://doi.org/10.1037/0022-0663.99.2.311.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6.

De Beni, R., Palladino, P., Pazzaglia, F., & Cornoldi, C. (1998). Increases in intrusion errors and working memory deficit of poor comprehenders. *The Quarterly Journal of Experimental Psychology Section A, 51*(2), 305–320. https://doi.org/10.1080/713755761.

de Milliano, I., van Gelderen, A., & Sleegers, P. (2016). Types and sequences of self-regulated reading of low-achieving adolescents in relation to reading task achievement. *Journal of Research in Reading, 39*(2), 229–252. https://doi.org/10.1111/1467-9817.12037.

DelVecchio, D. S., Jae, H., & Ferguson, J. L. (2019). Consumer aliteracy. *Psychology & Marketing, 36*(2), 89–101. https://doi.org/10.1002/mar.21160.

Dowse, R., & Ehlers, M. (2005). Medicine labels incorporating pictograms: Do they influence understanding and adherence? *Patient Education and Counseling, 58*(1), 63–70. https://doi.org/10.1016/j.pec.2004.06.012.

Egan, K. L., Schneider, M. C., & Ferrara, S. (2012). Performance level descriptors: History, practice and a proposed framework. In G. J. Cizek (Ed.), *Setting performance standards foundations, methods, and innovations* (2nd ed., pp. 79–106). New York, NY: Routledge.

Embretson, S. E., & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement, 11*(2), 175–193. https://doi.org/10.1177/014662168701100207.

Eme, E. (2011). Cognitive and psycholinguistic skills of adults who are functionally illiterate: Current state of research and implications for adult education. *Applied Cognitive Psychology, 25*(5), 753–762. https://doi.org/10.1002/acp.1746.

Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online, 5*(2), 50–79.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 430–445. https://doi.org/10.1037/0278-7393.16.3.430.

Grotlüschen, A., & Riekmann, W. (Eds.). (2012). *Funktionaler Analphabetismus in Deutschland: Ergebnisse der ersten leo. – Level-One Studie [Functional illiteracy in Germany: Results of the first Leo. – Level-one study]*. Münster: Waxmann.

Haberkorn, K., Pohl, S., Hardt, K., & Wiegand, E. (2012). *NEPS Technical Report for reading – Scaling results of starting Cohort 4 in ninth grade*. NEPS Working Paper no. 16. Bamberg: University of Bamberg, National Educational Panel Study. Retrieved 24 March 2020 from https://www.neps-data.de/Portals/0/Working%20Papers/WP_XVI.pdf.

Hahnel, C., Goldhammer, F., Kröhne, U., & Naumann, J. (2018). The role of reading skills in the evaluation of online information gathered from search engine environments. *Computers in Human Behavior, 78,* 223–234. https://doi.org/10.1016/j.chb.2017.10.004.

Hardt, K., Pohl, S., Haberkorn, K., & Wiegand, E. (2013). *NEPS Technical Report for reading – Scaling results of starting cohort 6 for adults in main study 2010/11*. NEPS Working Paper no. 25. Bamberg: University of Bamberg, National Educational Panel Study. Retrieved 24 March 2020 from https://www.neps-data.de/Portals/0/Working%20Papers/WP_XXV.pdf.

Hartig, J., & Frey, A. (2012). Konstruktvalidierung und Skalenbeschreibung in der Kompetenzdiagnostik durch die Vorhersage von Aufgabenschwierigkeiten [Construct validation and scale description in competence diagnostics through task difficulty forecasts]. *Psychologische Rundschau, 63*(1), 43–49. https://doi.org/10.1026/0033-3042/a000109.

Hartig, J., & Riekmann, W. (2012). Bestimmung der Level-Grenzen in der leo. – Level-One Studie. In A. Grotlüschen & W. Riekmann (Eds), *Funktionaler Analphabetismus in Deutschland: Ergebnisse der ersten leo. – Level-One Studie* [Functional illiteracy in Germany: Results of the first Leo. – Level-one study] (pp. 106–121). Münster: Waxmann.

Helder, A., Van Leijenhorst, L., & van den Broek, P. (2016). Coherence monitoring by good and poor comprehenders in elementary school: Comparing offline and online measures. *Learning and Individual Differences, 48,* 17–23. https://doi.org/10.1016/j.lindif.2016.02.008.

Karantonis, A., & Sireci, S. G. (2006). The bookmark standard-setting method: A literature review. *Educational Measurement: Issues and Practice, 25*(1), 4–12. https://doi.org/10.1111/j.1745-3992.2006.00047.x.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363.

Kirsch, I. S. (2001). The International Adult Literacy Survey (IALS): Understanding what was measured. *ETS Research Report Series, 2001*(2), i-61. https://doi.org/10.1002/j.2333-8504.2001.tb01867.x.

Kirsch, I. S., & Mosenthal, P. B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25*(1), 5–30. https://doi.org/10.2307/747985.

Koller, I., Haberkorn, K., & Rohm, T. (2014). *NEPS Technical Report for reading – Scaling results of Starting Cohort 6 for adults in main study 2012*. NEPS Working Paper no. 48. Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. Retrieved 24 March 2020 from https://www.neps-data.de/Portals/0/Working%20Papers/WP_XLVIII.pdf.

Landi, N. (2010). An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. *Reading and Writing, 23*(6), 701–717. https://doi.org/10.1007/s11145-009-9180-z.

Long, D. L., & Chong, J. L. (2001). Comprehension skill and global coherence: A paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1424–1429. https://doi.org/10.1080/01638539909545064.

Long, D. L., Seely, M. R., & Oppy, B. J. (1999). The strategic nature of less skilled readers' suppression problems. *Discourse Processes, 27*(3), 281–302. https://doi.org/10.1080/01638539909545064.

Lumley, T., Routitsky, A., Mendelovits, J., & Ramalingam, D. (2012). A framework for predicting item difficulty in reading tests. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Vancouver, 13–17 April 2012. Retrieved 24 March 2020 from http://research.acer.edu.au/pisa/5.

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction, 21*(3), 251–283. https://doi.org/10.1207/S1532690XCI2103_02.

McKoon, G., & Ratcliff, R. (2018). Adults with poor reading skills, older adults, and college students: The meanings they understand during reading using a diffusion model analysis. *Journal of Memory and Language, 102,* 115–129. https://doi.org/10.1016/j.jml.2018.05.005.

Mellard, D. F., & Fall, E. (2012). Component model of reading comprehension for adult education participants. *Learning Disability Quarterly, 35*(1), 10–23. https://doi.org/10.1177/0731948711429197.

Mitzel, H. C., Lewis, D. M., & Green, D. R. (2012). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards foundations, methods, and innovations* (pp. 249–282). New York, NY: Routledge.

Nusser, L., & Weinert, S. (2017). Appropriate test-taking instructions for students with special educational needs. *Journal of Cognitive Education and Psychology, 16*(3), 227–240. https://doi.org/10.1891/1945-8959.16.3.227.

OECD (Organisation for Economic Co-operation and Development). (2012). *PISA 2009 Technical Report*. Paris: OECD Publishing. https://doi.org/10.1787/9789264167872-en.

OECD (2013). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris: OECD Publishing. https://doi.org/10.1787/9789264204256-en.

Olsen, R.V., & Nilsen, T. (2017). Standard setting in PISA and TIMSS and how these procedures can be used nationally. In S. Blömeke & J.E. Gustafsson (Eds), *Standard setting in education* (pp. 69–84). Methodology of Educational Measurement and Assessment series. Cham: Springer. https://doi.org/10.1007/978-3-319-50856-6_5

Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods, 40*(4), 1001–1015. https://doi.org/10.3758/BRM.40.4.1001.

Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Amsterdam: Benjamins.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687.

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

Rouet, J.-F., Vidal-Abarca, E., Erboul, A. B., & Millogo, V. (2001). Effects of information search tasks on the comprehension of instructional text. *Discourse Processes, 31*(2), 163–186. https://doi.org/10.1207/S15326950DP3102_03.

Sonnenleitner, P. (2008). Using the LLTM to evaluate an item-generating system for reading comprehension. *Psychology Science Quarterly, 50*(3), 345–362.

Tighe, E. L., & Schatschneider, C. (2016). Examining the relationships of component reading skills to reading comprehension in struggling adult readers: A meta-analysis. *Journal of Learning Disabilities, 49*(4), 395–409. https://doi.org/10.1177/0022219414555415.

Todaro, S., Millis, K., & Dandotkar, S. (2010). The impact of semantic and causal relatedness and reading skill on standards of coherence. *Discourse Processes, 47*(5), 421–446. https://doi.org/10.1080/01638530903253825.

UN (United Nations) (2016). *Sustainable Development Goal 4: Targets and indicators* [dedicated webpage]. New York: UN. Retrieved 26 March 2020 from https://sustainabledevelopment.un.org/sdg4#targets .

Vágvölgyi, R., Coldea, A., Dresler, T., Schrader, J., & Nuerk, H.C. (2016). A review about functional illiteracy: Definition, cognitive, linguistic, and numerical aspects. *Frontiers in Psychology, 7*, Art. 1617. https://doi.org/10.3389/fpsyg.2016.01617

van Beusekom, M. M., Grootens-Wiegers, P., Bos, M. J. W., Guchelaar, H.-J., & van den Broek, J. M. (2016). Low literacy and written drug information: Information-seeking, leaflet evaluation and preferences, and roles for images. *International Journal of Clinical Pharmacy, 38,* 1372–1379. https://doi.org/10.1007/s11096-016-0376-4.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450. https://doi.org/10.1007/BF02294627.

Weeks, L. (2001, April 5). The no-book report: Skim it and weep. *Washington Post*, 14 May 2001 [online article]. Retrieved 24 March 2020 from https://www.washingtonpost.com/archive/lifestyle/2001/05/14/the-no-book-report-skim-it-and-weep/590f488d-c1fd-498d-81da-55269a967e94/.

White, S. (2012). Mining the text: 34 text features that can ease or obstruct text comprehension and use. *Literacy Research and Instruction, 51*(2), 143–164. https://doi.org/10.1080/19388071.2011.553023.

Wicht, A., Durda, T., Krejcik, L., Artelt, C., Grotlüschen, A., Rammstedt, B., & Lechner, C.M. (forthcoming). Risk and protective factors for ascents from or descents into low reading literacy in adulthood: Insights from the German National Educational Panel Study. Manuscript submitted for publication.

Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2005). Less skilled comprehenders' ERPs show sluggish word-to-text integration processes. *Written Language & Literacy, 8*(2), 157–181. https://doi.org/10.1075/wll.8.2.10yan.

Zimmermann, S. (2016). *Entwicklung einer computerbasierten Schwierigkeitsabschätzung von Leseverstehensaufgaben* [Development of a computer-based difficulty estimation of reading comprehension tasks]. NEPS Working Paper no. 64. Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. Retrieved 24 March 2020 from https://www.neps-data.de/Portals/0/Working%20Papers/WP_LXIV.pdf.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Tabea Durda is a PhD candidate at the University of Bamberg, Germany, and a research fellow at the Leibniz Institute for Educational Trajectories. Her research focuses on assessment of low reading proficiency and methodological challenges in proficiency scaling, impact of institutional learning settings on language and reading development from early childhood to adulthood.

Cordula Artelt is professor for Longitudinal Educational Research at the University of Bamberg, Germany, and the director of the Leibniz Institute for Educational Trajectories. Her research focuses on reading and text comprehension, metacognition and self-regulated learning, methods of large-scale assessment, assessment of competencies and teacher judgements.

Clemens M. Lechner is the head of the Scale Development and Documentation team in the Department of Survey Design and Methodology at GESIS – Leibniz-Institute for the Social Sciences in Mannheim, Germany. His research focuses on individual differences in competence development, cognitive abilities, personality, and values.

Beatrice Rammstedt is professor of Psychological Assessment, Survey Design and Methodology at the University of Mannheim, Germany, as well as vice president and scientific director of the Survey, Design and Methodology department at GESIS – Leibniz Institute for the Social Sciences. Her research focus includes methodological topics like questionnaire design and validation and methodology of cross-cultural comparative large-scale studies as well as content-oriented topics like the relevance of socio-emotional skills for school and life success.

Alexandra Wicht is a senior researcher at the University of Siegen, Germany. Her research focuses on social inequality, educational and labour market sociology, including competency development, aspirations, school-to-work transitions, and effects of social contexts on individual outcomes.

Affiliations

Tabea Durda[1] · Cordula Artelt[1] · Clemens M. Lechner[2] · Beatrice Rammstedt[2] · Alexandra Wicht[3]

> Cordula Artelt
> cordula.artelt@lifbi.de

> Clemens M. Lechner
> clemens.lechner@gesis.org

> Beatrice Rammstedt
> beatrice.rammstedt@gesis.org

> Alexandra Wicht
> alexandra.wicht@uni-siegen.de

[1] LIfBi – Leibniz Institute for Educational Trajectories, Bamberg, Germany

[2] GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany

[3] University of Siegen, Siegen, Germany