Gehrer, Karin; Zimmermann, Stefan; Artelt, Cordula; Weinert, Sabine

NEPS framework for assessing reading competence and results from an adult pilot study

Journal for educational research online 5 (2013) 2, S. 50-79

urn:nbn:de:0111-opus-84247

Journal for Educational Research Online
Journal für Bildungsforschung Online
Volume 5 (2013), No. 2, 50–79

Karin Gehrer, Stefan Zimmermann, Cordula Artelt
& Sabine Weinert

NEPS framework for assessing reading competence *and* results from an adult pilot study

Abstract

This article sketches the framework for assessing reading competence across the lifespan in the German National Educational Panel Study (NEPS). It gives a detailed presentation of the two central dimensions in the framework: (a) text functions and text types and (b) the cognitive requirements of reading tasks. These are discussed against the background of relevant theoretical models and research findings. A pilot study of 447 adults is reported that analyzed the dimensionality and difficulty of the reading competence test for adults. Results indicated that the test meets the NEPS research goals. The article focuses particularly on whether text types and cognitive requirements prove to be appropriate structural elements for the framework, that is, whether each distinguishes sufficiently between good and poor readers. Results also report on the dimensionality of the reading competence test. A comparison between one unidimensional and two different multidimensional models examined whether the text types and/or cognitive requirements of the items/tasks are separate dimensions of reading competence, or whether the reading competence measured in NEPS can – as intended – be conceived as a unidimensional construct. Results are discussed against the background of the scope and limitations of the NEPS reading competence test and the relevant research literature on reading competence.

Karin Gehrer, lic.phil., Dipl. HLA (corresponding author) · Stefan Zimmermann, M.Sc., National Educational Panel Study (NEPS), University of Bamberg, Wilhelmsplatz 3, 96045 Bamberg, Germany
e-mail: karin.gehrer@uni-bamberg.de
stefan.zimmermann@uni-bamberg.de

Prof. Dr. Cordula Artelt, Chair of Educational Research, University of Bamberg, Markusstraße 3, 96045 Bamberg, Germany
e-mail: cordula.artelt@uni-bamberg.de

Prof. Dr. Sabine Weinert, Chair of Psychology I: Developmental Psychology, University of Bamberg, Markusplatz 3, 96045 Bamberg, Germany
e-mail: sabine.weinert@uni-bamberg.de

Keywords
Reading competence; Language assessment; Tests; Dimensionality; Cognitive requirements; Text functions

NEPS-Rahmenkonzeption zur Messung von Lesekompetenz *und* Resultate einer Pilotstudie mit Erwachsenen

Zusammenfassung
Im Rahmen des Artikels wird das Rahmenkonzept zur Messung von Lesekompetenz über die Lebensspanne im Nationalen Bildungspanel (NEPS) skizziert. Dabei werden zwei zentrale Dimensionen dieser Rahmenkonzeption, (a) Textfunktionen bzw. Textsorten und (b) kognitive Anforderung der Leseitems, im Detail dargestellt und vor dem Hintergrund relevanter theoretischer Modelle und Forschungsbefunde diskutiert. Zudem werden Ergebnisse einer Pilotstudie berichtet, die die Angemessenheit des Lesekompetenztests für die Forschungsintentionen des NEPS basierend auf Analysen zur Dimensionalität und Schwierigkeit der Lesekompetenztests auf Basis einer Stichprobe von 447 Erwachsenen darstellen. Spezieller Fokus des Beitrags ist dabei die Frage, ob Textsorten und kognitive Anforderungen als strukturelle Elemente der Rahmenkonzeption insofern angemessen sind, als dass beide (auch) im Erwachsenenalter erlauben, hinreichend zwischen guten und schwachen Lesern zu differenzieren. Zudem werden Ergebnisse zur Dimensionalität des Lesekompetenztests dargestellt. Basierend auf einem Vergleich eines eindimensionalen und zwei unterschiedlichen mehrdimensionalen Modellen gehen wir der Frage nach, ob Textsorten und/oder kognitive Anforderungen der Items separate Dimensionen der Lesekompetenz ausmachen oder die im NEPS gemessene Lesekompetenz – wie intendiert – als eindimensionales Fähigkeitskonstrukt aufgefasst werden kann. Die Ergebnisse werden vor dem Hintergrund der Möglichkeiten und Limitationen des Lesekompetenztests des Nationalen Bildungspanels und der relevanten Forschungsliteratur zur Lesekompetenz diskutiert.

Schlagworte
Lesekompetenz; Kompetenzmessung; Test; Dimensionalität; Kognitive Anforderungen; Textsorten

1. Introduction

Being able to read is a key competence for coping with the demands of everyday life and participating in society. It also remains crucial throughout life for the acquisition and further development of knowledge and skills in countless fields. However, longitudinal studies on the development and the role of reading compe-

tence beyond school age have been either very infrequent or have covered relatively short timespans. The German National Educational Panel Study (NEPS) aims to track reading competence coherently across long stretches of the lifespan, thereby providing empirical access to one of its central issues (see Blossfeld, Roßbach, & von Maurice, 2011). However, such a consistent and coherent longitudinal empirical assessment raises both content-related and methodological challenges for both the framework of competence measurement in NEPS and the development of appropriate instruments (see Weinert et al., 2011). One central starting point for the NEPS framework is an orientation toward the functionality and everyday relevance of the competencies studied. This orientation draws on the concept of *literacy* in international comparative studies with a focus on enabling participation in society (see OECD, 1999).

The article starts by explaining how the current literature understands reading competence and text comprehension as an active process of construction occurring on several levels. It then presents theoretical and pragmatic considerations that take account of earlier concepts and studies of reading competence within the framework of large-scale assessments (LSAs) and form the background to specifying the NEPS framework for measuring reading competence. It reports on the decision not to use discontinuous texts, the state of research on text typology, and the selection of concrete text types in other LSAs; discusses work on the cognitive requirements of text comprehension tasks; and, finally, explains the most important dimensions (text types, cognitive requirements, task formats) of the NEPS framework for measuring reading competence. It also discusses the standards of test development within NEPS and how these are applied in the development of instruments and empirical pilot studies.

Finally, a larger pilot study with adults is used to analyze the parameters for item difficulty and present analyses of the dimensionality of the reading competence test by contrasting the hypothesized unidimensional model with multidimensional models of text functions or cognitive requirements as independent dimensions.

2. State of research

It was only during the phase of (radical) constructivism that research and models of text comprehension found their way back to the old hermeneutic insight that "the 'text' is finally something that constitutes itself in the experience of the recipient" (Hess-Lüttich, 1996, p. 7, translated). Since the work of Kintsch and van Dijk (1978) and Kintsch (1998; van Dijk & Kintsch, 1983), research in cognitive psychology has viewed the reading of texts as a complex active process requiring a number of interacting subabilities (see, for overviews, Artelt et al., 2005; Christmann & Groeben, 1999; Richter & Christmann, 2002). This leads to a distinction between processes on a lower and a higher hierarchic level. The lower level of the process

hierarchy contains the necessary automatic substeps of perception, identification, analysis, and all the decoding processes that lead to word recognition. On the level of sentences, the single semantic statements, the so-called propositions, are assessed within the syntactic sentence structure and related to each other within the process of local coherence formation in order to interpret a phrase meaningfully. On higher hierarchical process levels, there are also cognitive-active processes of selection, construction, and integration in which whole sequences of propositions are linked together, consolidated, selected, and generalized so that it becomes possible to understand complete text elements on the text level. Finally, comprehensive processes of global coherence formation produce so-called macrostructures on a high level of abstraction in order to grasp the global gist of a text.

Research based on the work of Kintsch (1970/1982) and Kintsch and van Dijk (1978) basically follows what was originally Bartlett's schema theory by assuming that the reader applies a cognitive-active comprehension process and uses the representation of the text content presented to finally build up a mental model (in line with Collins, Brown, & Larkin, 1980, as cited in Quathamer, 1998, p. 16). The quality and process of the mental representation (mental model), however, depend, among others, on the reader's individual abilities and capacities. In the process of building up a coherent mental representation, the reader's structural and content-specific knowledge base as well as his or her general knowledge of the world play a special role, because, for example, knowing about the function and particular structure of a special kind of text (be it a fairy tale, an entry in a dictionary, a newspaper article, or whatever) facilitates its reception. Moreover, prior content knowledge makes it easier in general to form a cognitive text representation, because both linking together associated concepts and drawing conclusions on what one has read depend on what one already knows (Kintsch, 1998; Schnotz, 1988).

3. Theoretical and pragmatic prior considerations for the NEPS framework for assessing reading competence

3.1 General considerations on the formation of models in NEPS

In contrast to models in cognitive psychology such as that described by Kintsch and van Dijk (1978) with its focus on the internal processes of information retrieval from the text and a small-scale and multidimensional process analysis, any models focusing on output have to abstract the measurement of performance from the internal structure and process analysis (Schnotz & Dutke, 2004). However, even in an LSA, the goal has to be to achieve the best possible agreement between the findings of cognitive research and the models constructed for test development.

Previous reading competence tests in the frameworks of LSAs have chosen different conceptual focuses. These range from a strong orientation toward the idea of

literacy in international studies of reading competence – such as the International Adult Literacy Survey (IALS; e.g., OECD & Statistics Canada, 1995) in the 1990s or the multicycle comparisons of school performance in the Programme for International Student Assessment (PISA; OECD, 1999, 2001, 2004, 2007, 2009, 2010; see, for Germany, e.g., Artelt, Schiefele, Schneider, & Stanat, 2002; Baumert et al., 2001; Klieme et al., 2010) – to projects based more strongly on linguistic models such as the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001) that provides an orientation for studies such as the Level-One Survey (leo) of functional illiteracy (Grotlüschen & Riekmann, 2011), the Workforce Literacy Development Study (lea) on the literacy needs of employees (Grotlüschen, Kretschmann, Quante-Brandt, & Wolf, 2011), or the national reading competence test for secondary school graduates in Switzerland (EVAMAR; Eberle et al., 2008).

When operationalizing reading competence with stimulus texts and items, one general approach is to categorize the underlying texts according to the type of situation in which they are applied. In the field of reading, this has commonly led to a focus on the reasons for reading. For example, the CEFR defines the requirements on learners in terms of communication situations in (a) education, (b) work, (c) the personal domain, and (d) the public domain (Council of Europe, 2001, p. 45). The framework for reading competence in the PISA study also uses comparable situations (Baumert, Stanat, & Demmrich, 2001, p. 24; OECD, 1999).

When designing the longitudinal assessment of reading competence from childhood to retirement in NEPS, these four situations do not offer any coherence across the lifespan – the work domain is lacking in childhood and school age, and not all adults continue to take advantage of education. Therefore, NEPS works with a concept that is oriented less toward the reasons for reading and the reading situations associated with this, but focuses predominantly on the functions of text along with the types of text associated with these functions and how these relate to the cognitive requirements of reading.

3.2 Text functions and text types

Cognitive and psycholinguistic research literature addresses the ways in which the structural features and different communicative functions of text types or genres influence text comprehension because it is assumed that readers form multiple mental representations. These include constructing not only mental representations of the surface text (lexis, syntax), the semantic content, and the situation described, but also a mental representation of the communicative intention of the author and of the text genre (Schnotz & Duthke, 2004, p. 73). The following section will explain the choice of text types used in the NEPS reading framework.

3.2.1 Continuous versus discontinuous texts

One class of text that has become of increasing interest to research in recent years is that of discontinuous (or also noncontinuous) texts. For example, this class plays a major role in the design of PISA (e.g., Artelt, Stanat, Schneider, & Schiefele, 2001; Schaffner, Schiefele, Drechsel, & Artelt, 2004).

Continuous texts are the classic "running texts" of prose (in contrast to classic verse; see Göttsche, 2010, p. 38) that transport exclusively verbal information in the form of letters. Noncontinuous or discontinuous texts extend this by linking the written verbal information to pictorial information in "logical images" (tables, graphs, diagrams) or "realistic" images (illustrations) that are applied as functional rather than decorative elements. The combination of continuous and discontinuous texts results in a broader concept of reading competence (see, for PISA, Baumert et al., 2001; Schnotz & Dutke, 2004, p. 63). However, such a broader concept of reading competence has been criticized (see, e.g., Beck & Klieme, 2007a; Wieler, 2003), because of the suspected and in part also empirically confirmed heterogeneity of the demands imposed on the reader (Artelt et al., 2001; Artelt & Schlagmüller, 2004; Artelt, Stanat, Schneider, Schiefele, & Lehmann, 2004; Baumert et al., 2002; Schnotz & Dutke, 2004).

Whereas Mayer (1997, as cited in Schnotz & Bannert, 1999, p. 3) assumes that the processes of text comprehension and image comprehension run parallel, and contrasts the propositional representation of the text with the imaginal representation of the image, Schnotz and colleagues emphasize that the combination of verbal and pictorial messages in discontinuous texts leads to a new mental model construction when reading. As a result, this type of process can be assessed only with an integrative model of text and image comprehension (e.g., Schnotz & Dutke, 2004; see, for detailed empirical work, Schnotz & Bannert, 1999).

Because of the additional requirements of longitudinal modeling, the NEPS framework and the test construction based on it concentrate on a more homogeneous understanding of reading competence and focus exclusively on the classic continuous forms of text. Therefore, pictorial elements (drawings, clip arts, photographs) are used very sparingly in the NEPS test booklets and only for decoration. The mostly humorous and realistic illustrations are only there to motivate participants. In other words, participants do not have to construct any systematic relations between the (decorative) pictorial elements and the written verbal information. So-called "logical" images such as diagrams and extensive tables from which auxiliary or even the main information has to be linked to the written continuous text have been excluded consistently from the model formation and test development.

As pointed out above, the decision to not use discontinuous texts to measure reading competence in NEPS is based on the argument that continuous and discontinuous forms of text to some extent require different types of reading and comprehension processes. This argument has also been confirmed indirectly through dimension analyses (Artelt & Schlagmüller, 2004). The focus on continuous texts

is also supported by the justified supposition that written continuous text will continue to be the main medium despite the marked growth in the use of discontinuous texts and logical image elements in modern information societies (Schnotz & Dutke, 2004, p. 63). Admittedly, the gain in the precision of measurement through this concentration on purely continuous forms of text does have its price: Limiting the concept of reading competence in this way does reduce some of the relevance to daily life, when it is considered that the functionality concept of an extended reading competence in today's and future media societies naturally also includes the comprehension of functional picture elements.

3.2.2 Continuous text types

Within the continuous forms of text chosen for NEPS, there has always been a dichotomization taken from literary studies and extended by text typology research between literature (of the epic genre, i.e., prose) and nonliterature prose. The distinction between fictional and nonfictional texts (Scheffel, 2010a; Werlich, 1975) – the latter also being called factual texts – also points to the difference between literary texts and pragmatic texts (e.g., Abraham, 2003) or so-called functional texts (Brinker, 1985). Beyond this basic dichotomy[1], neither the form criticism of German studies of comparative literary studies nor the research on text types in textual linguistics reveal any general homogeneity of text typology (see, for overviews, Adamzik, 2010; Zymner, 2010). Hence, we cannot fall back on any broadly accepted classification of text types or text classes when developing reading competence tests.

Nonetheless, Brinker's linguistic definition shaped by action theory in the 1980s can serve as a "standard definition" of text type (Adamzik, 2010, p. 96): "Text types are conventionally valid patterns for complex verbal actions and can be described as specifically typical ties between contextual (situational), communicative-functional, and structural (grammatical and thematic) features" (Brinker, 1985, p. 124, 2010, p. 135, translated). Hence, the function of a text can be seen as a basic criterion for differentiating text types (Brinker, 1983, pp. 144–147). Therefore, the text type is by its very definition always tied to the dominant communicative function that determines it (Brinker, 1985, p. 128). Often, texts fulfill more than one communicative function or speech acts; since Searle's speech act theory a text "is defined as a *complex* verbal action i.e. as a hierarchically structured composition of speech acts, of those one dominates the others" (Brinker, 1983, p. 136, translated)[2].

1 The categories originating in Aristotle of fictional ("poetry") versus factual ("historiography") cannot always be separated strictly, because they overlap in certain text types such as autobiographies or the lyrical ego of a poem (Scheffel, 2010a, p. 30).

2 Brinker points out that the pure quantity of some types of semantic sentences cannot be an indicator for baseline dominance; finally, the decision on the dominance of one function in a text is given by the context or situation (Brinker, 1983, p. 135).

When attempting to find some structure for the incredible variety of text types[3], most early classification approaches in textual linguistics designed to differentiate according to the function of a text referred to the three basic text functions assumed in Bühler's (1934) Organon communication model: (a) expressive function, (b) referential function, and (c) persuasion function. In the 1980s, Brinker extended this model to five basic textual functions within instruction texts: (a) informative (in text types such as news reports, descriptions, and textbooks), (b) persuasive, (c) obligatory, (d) contacting, and (e) declarative.

The NEPS framework follows Brinker (1983, 1985, 2010) in defining text types according to their function, but does not stick to his typology of five text classes of factual or pragmatic texts[4] (see framework section).

3.2.3 Text types in other LSAs

Analog to the inconsistent state of linguistic theory, the major studies on reading competence also apply a variety of operationalizations. For example, the reading test in the longitudinal Assessment of Student Achievements in German and English as a Foreign Language (DESI) used the traditional dichotomy of text types and combined literary texts with information texts, applying two of each at every measurement time (DESI Konsortium, 2006, p. 4). PISA draws on a greater variety of texts, and its original framework is based on the text type model of Mosenthal and Kirsch (1991) that distinguishes between descriptive, narrative, expositional, and argumentative types within continuous texts. The actual design of the PISA reading test reveals six text types: descriptive, expositional, argumentative, directive (or also instructional), documenting (or records), and "narrative" – based on the five-part text typology developed by Werlich (1975, 1976).[5] The fourth cycle of PISA (2009) supplemented paper-based reading competence tests with an optional computer-based test on reading electronic texts. The framework of reading competence was correspondingly extended to include electronic texts, and, among others, the subscales assessing cognitive requirements were adapted to the requirements of electronic text types. This also led to the replacement of the original text

3 Referring to the text type concepts in daily language, one can already find more than 1,600 terms in the German Duden Spelling Dictionaries of the 1970s. About 500 of these are basic text types such as Bericht [report]; the others are secondary compounds such as *Reisebericht* [travel report], *Ergebnisbericht* [outcome report], or *Wetterbericht* [weather report] (Dimter, 1973, as cited in Brinker, 1985, p. 121).

4 Brinker's typology derives five basic text classes from the basic functions reported above that are each tied to a corresponding text function. These are information texts, persuasion texts (advertisement, newspaper commentary, instructions, directions for use, law, sermon, etc.), obligatory texts (contract, pledge, tender, etc.), contact texts (with text types such as thank you letter, love letter, postcard, etc.), and declarative texts (mandates, deeds, wills, etc.) (Brinker, 1985, p. 125).

5 PISA uses "narrative" in a very general sense that also applies to instruction or information prose compared to the narrative as genre in literary studies (see, on the theory of the narrative, Scheffel, 2010b, p. 329).

type "documentary" with texts serving the function "transactional (exchange of information)" (Naumann, Artelt, Schneider, & Stanat, 2010).

3.3 Cognitive requirements of reading tasks in other LSAs

In line with the given paradigm of viewing reading competence as an active comprehension process in which many cognitive steps impose different cognitive demands, LSAs of reading competence tap the various facets of reading competence or the cognitive requirements in the process of text comprehension with different kinds of tasks.

The most comprehensive international study of adult reading competence to date, the International Adult Literacy Survey (IALS/ALL; e.g., OECD & Statistics Canada, 1995) based its measurement of reading competence in the 1990s on the literacy model of Kirsch and Mosenthal (Kirsch, Jungeblut, & Mosenthal, 1998; Mosenthal & Kirsch, 1991; Mosenthal, 1996) and studied several facets of literacy: (a) prose literacy: the comprehension of flow texts with or without pictorial information; (b) document literacy: the reading and comprehension of documents such as forms, timetables and other tables, diagrams, and illustrations; and (c) quantitative literacy: the gathering of numerical information from forms, tables, and other texts that also involves drawing mathematical conclusions (OECD & Statistics Canada, 1995, p. 14). In the IALS/ALL, the cognitive requirements of items were distinguished according to three main aspects of information processing: locating, integrating, and generating. Locating deals with taking information from text, which partly involves drawing conclusions. Integrating requires the reader to piece together information from two or more locations in the text; these can either lie within one section or be distributed across several sections. Generating requires the reader to further process information in the text (e.g., to deliver a written answer) and to draw text-based conclusions, at times, also on the basis of background knowledge (Kirsch, 1995, p. 30; see also Kirsch, 2001). Kirsch (2001) and Mosenthal and Kirsch (1998) developed an additive item rating scheme for this that considered both features of the task or item and the interaction with the text needed to solve the text comprehension task. Depending on which different cognitive processes are necessary for the task processing (difficulty-generating features), the rating scheme assumes the item to be easier or more difficult. In various studies, the authors were able to show that estimates of task difficulty based on their rating scheme were powerful predictors of empirically ascertained task difficulties. They distinguished between the following three factors: (a) type of match, (b) type of information requested, and (c) distracting information. The processes making up the first factor, type of match, are locating, circulating, integrating, and generating. According to the second factor, type of information requested, questions on abstract information are more difficult to answer than questions on concrete things. The third factor, distracting information, addresses the plausibility of incorrect options in the text. An item is easier when there is no distracting information

in the text and more difficult when distractors are located in the same paragraph as the answer to be sought.[6] The systematization of cognitive requirements developed within the framework of the IALS study also forms the basis for the differentiation into three subscales (finding information in the text, drawing text-related conclusions, and reflecting and assessing) in the PISA framework (see below; see also OECD, 1999, 2009) that are each represented by tasks of varying difficulty (assessed empirically or in part with reference to the above schema).

The international elementary school reading study, Progress in International Reading Literacy Study (PIRLS; e.g., Mullis, Martin, González, & Kennedy, 2003; for the German Internationale Grundschul-Lese-Untersuchung, IGLU; e.g., Bos et al., 2003, 2004) used a framework distinguishing between four comprehension processes that were also reflected in the item construction: (a) recognizing and reporting explicitly given information; (b) drawing simple conclusions; (c) drawing complex conclusions, justifying them, and interpreting what one has read; and (d) testing and assessing language, content, and text elements. The first two tasks require the use of information inherent in the text; conclusions are reached by forging relations between given parts and sections. In the more complex comprehension processes (c) and (d), in which respondents have to reflect on content or structures, it is also necessary to draw on external knowledge (Bos et al., 2003, pp. 76–77). The DESI study (Beck & Klieme, 2003, 2007b; Nold & Willenberg, 2007; Willenberg, 2007) oriented its measurement of reading competence in 9th-grade classes toward text research within cognitive psychology by Kintsch[7] (1994) or van Dijk and Kintsch (1983). It used a process model of reading competence with six requirements on the theoretical levels of (a) information, (b) inferences, (c) focusing, (d) knowledge, (e) links, and (f) mental model (Nold & Willenberg, 2007; Willenberg, 2007; see, for a criticism, e.g., Bremerich-Vos & Grotjahn, 2007). Their empirical findings revealed that these could be aggregated to form the following four cognitive requirements: (a) finding information on the level of words in the context of the sentence; (b) local reading: drawing conclusions (inferences) and abstracting the topics in single sections; (c) forging links between different paragraphs and text passages – also with recourse to one's own prior knowledge; and (d) mental model: integrating the central aspects of the text (see Garbe, Holle, & Jesch, 2009, p. 27).

The international student assessment program PISA uses the theoretical structure of reading competence to differentiate five cognitive requirements: (a) developing a general understanding of the text; (b) gathering information; (c) developing a text-related interpretation; (d) reflecting on the content of the text; and (e) reflecting on the form of the text. Three subdimensions could be confirmed empir-

6 According to the schema, other factors influencing task difficulty are the number of sentences within which the correct answer is to be found, the number of answers sought and whether this number is reported, how far the information given in the question matches the information in the text, and how far the answer can be taken from the text or has to be constructed by the respondent (see Kirsch, 2001).

7 Whereas Kintsch talks about the "situation model" of the text, DESI adopts the "mental model" concept from Christmann and Groeben (1999) for the highest level.

ically: gathering information (b), text-related interpreting (a and c), and reflecting and assessing (d and e). These were added to the total scale of reading competence as report scales (Artelt et al., 2001).

NEPS is oriented toward the aforementioned theoretical analyses and empirical findings from various LSAs, and applies the distinction between cognitive requirements and difficulty-generating features for the framework and operationalization of reading competence.

4. The NEPS reading competence framework

Reading makes it possible to access and acquire a variety of life and knowledge domains. The range of reasons for reading is very broad, and reading simultaneously fulfills a multitude of different functions (see, e.g., Groeben & Hurrelmann, 2004). These range from the reading that is essential for further training and lifelong learning, across broadening one's general knowledge, up to literary and aesthetic reading. Texts convey not only information and facts but also ideas, values, and the contents of culture. The concept of reading competence in NEPS is based accordingly on a functional understanding of reading competence as also reflected in the literacy concept (see also OECD, 2009). The focus is on handling texts competently in various characteristic everyday situations.

The concept of reading competence in the NEPS framework concentrates on reading competence as text comprehension: It assesses the comprehension performance shown in replies to questions referring to the specific underlying text. Research focuses on the abilities to read a text and understand it appropriately – both as a whole and in its single statements. The emphasis is on understanding what is in the text and not primarily on memory performance for text material that has been read but is no longer available.

To represent the concept of reading competence as coherently as possible over the entire lifespan, the framework for the NEPS reading competence test specifies three features that have to be taken into account in each age- or stage-specific[8] form of the test: (a) text functions and text types, (b) cognitive requirements, and (c) item formats. However, in contrast to other LSAs that use a booklet design and thus are able to administer far more test items in the same time period, the items within each NEPS reading competence test are the same for each student within the particular age group/stage. As a result, the set of items administered is rather limited (ca. 30 items referring to five texts per measurement time). NEPS aims to assess reading competence as a comparatively homogeneous unidimensional construct. The features differentiated in the framework and test construction serve pri-

8 NEPS divides education trajectories into eight educational stages such as "From Kindergarten to Elementary School (Stage 2)". Some stages apply specific additional tests that are not continued longitudinally across the total lifespan. For example, Stage 2 is assessing phonological awareness (see Berendes, Weinert, Zimmermann, & Artelt, 2013, this issue).

marily to account systematically for the breadth of text types, cognitive requirements, and item formats.

4.1 Text functions and text types

NEPS distinguishes five text functions and their associated text types that are taken into account in each form of the test: (a) information texts, (b) commenting or argumenting texts, (c) literary texts, (d) instruction texts, and (e) advertising texts. This selection was based on the assumption that these five text functions are relevant for the everyday lives of participants of all different ages.

The continuous forms of text applied in NEPS can be characterized in terms of their functions or types (see, for more detail, Gehrer & Artelt, 2013): *Texts imparting information* represent basic texts for learning, the fundamental acquisition of knowledge, and finding information. Examples are articles, dispatches, reports, and announcements. Texts with a *commenting or argumenting function* take a particular stance or they query something, balance out arguments for and against, or include a reflective view (reader's letters, discussions, essays, academic papers). Texts with a *literary-aesthetic function* are short stories, excerpts from novels, or narratives. Special literary text types such as theater plays, satires, or poems are excluded because reading them probably depends strongly on types of education and curricula. The fourth category is made up of text types that convey *user guidance* such as assembly instructions, operating manuals, package insets for medication, work instructions, cooking recipes, and so forth. The fifth category *advertising* (*appeals, advertisements, announcements*) contains texts for advertising, job announcements, leisure activities, and so forth.[9]

The five selected text functions and their associated text types are operationalized in each test booklet as a longitudinal concept across the lifespan; that is, each test booklet measuring reading competence contains a total of five texts corresponding to these five text functions.

In contrast to PISA, NEPS is not applying any discontinuous texts with pictorial information from diagrams, tables, or graphical illustrations. Discontinuous texts are not included in the NEPS framework, as mentioned before, because they impose special requirements on readers.

9 For the reading test, preference was given to selecting texts that are as prototypical as possible for each text type. Nonetheless, the borders between some text types are not so fixed, for example, many instruction texts only report the actual operating instructions after an introductory section containing general information. There are also mixed types, for example, when advertising takes quotes from literature and integrates them humorously. According to Brinker (1983) we orientated ourselves on the all-dominant communicative function of the given text and tried to avoid such extremely mixed types when selecting texts for the NEPS reading test. For the type "information text" we set limits in the narrower sense to avoid an intermixture with other subtypes e.g. compulsory information texts (see Gehrer & Artelt, 2013, pp. 175–178).

In order to measure differences in text comprehension rather than differences in prior knowledge, the demands on the specific prior knowledge of the test persons should be kept as low as possible in a reading competence test (see, for an overview, e.g., Köster, 2003). Therefore, NEPS fundamentally excludes texts requiring specific prior knowledge from both its framework and the test construction. Such texts can be poems whose reception builds on prior knowledge of types of rhyme and verse, metric, and their place in literary history, but also specialized texts and those types of information texts with a discipline-specific vocabulary requiring special prior knowledge. Texts are primarily selected to have topics reflecting a general knowledge of the world. In addition, several different approaches were used to keep the demands on the reader's prior knowledge as low as possible, for example, by formulating the items so that they are closely related to the text (e.g., by asking which of several correct statements can be found in the text and which not) or by telling the reader to refer to the text when answering questions both in the instructions, after each text before answering the items, and, at times, even within the question stem.

4.2 Cognitive requirements

The second feature of the framework and thus of the task construction for measuring reading competence in NEPS is cognitive requirements. As mentioned above, various types of cognitive requirements can be derived from the cognitive psychological literature on reading competence and text comprehension on both hierarchically low (decoding, word recognition) and hierarchically high levels (local and global text comprehension) (e.g., Kintsch, 1998; Richter & Christmann, 2002). As also mentioned above, the latter have been operationalized slightly differently in various LSA studies of reading competence. The NEPS concept of reading competence reflects the higher cognitive requirements in three specific types of item. These variants are labeled types, because they are not based on any explicit assumptions that one type of item is necessarily more or less difficult than another type of item. However, each type of item taps another kind of cognitive requirement in the comprehension process. The first type is items on "finding information in text" (Type 1). These are items in which readers have to find detailed information on the sentence level, that is, to decode and recognize statements and propositions. A first version of this type of item is designed so that the formulation in the stimulus text and the item is identical (Type 1.1); in the second version, the two formulations differ (Type 1.2).

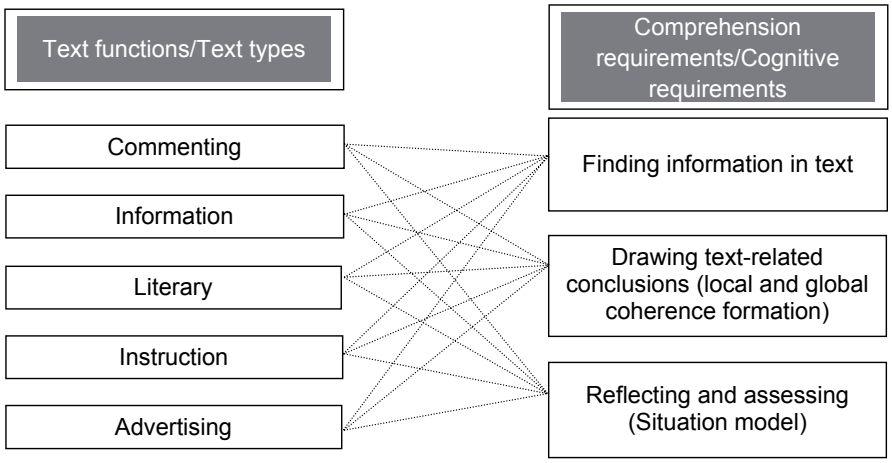
In Type 2 items, on "drawing text-related conclusions", inferences have to be drawn from several sentences in order to construct local or global coherence. In the first version, this has to be done with neighboring sentences; in the second version, with several sentences spread across whole sections of text; and in the third version (Type 2.3), it is necessary to understand the major ideas in the text, which requires the comprehension of larger or more complex sections of relevant text.

The third type of item includes the cognitive requirements of "reflecting and assessing". In the first version (Type 3.1), readers have to comprehend the central idea, event, or message in a text; in the second version (Type 3.2), they have to be able to recognize the purpose and intention of a text and judge its credibility. The third version integrates the need for background knowledge (Type 3.3). The requirements reflect, among others, the need to represent the text in the form of a situation model or a mental model.

Although, at first glance, the types of cognitive requirements may seem to be ranked according to difficulty, they differ far more qualitatively, and primarily reflect a broad spectrum of requirements on the text level within the framework of a reading competence test. In contrast to rather hierarchical models of cognitive processes on the word-, sentence- and text-level the NEPS framework and reading competence test focus exclusively on the cognitive requirements of the higher hierarchical levels in the construction processes of text comprehension. A text comprehension test like the NEPS reading competence test takes the lower processes for granted but does not assess them. The various types of tasks in the NEPS reading competence test manifest different kinds of complex cognitive requirements (finding information in text, drawing text-related conclusions, reflecting and assessing). By way of example, least of all in task type 1 a very complex process occurs while words, clause constituents, subordinate clauses and compound sentences must be combined, weighted and compared among themselves and against the stimulus text for verification or falsification.

For each type of requirement, items can be easy, intermediate, or difficult; hence, there are not only easy items on the level of reflecting but also difficult items on the level of finding information or drawing conclusions. More detail can be found in the empirical part of this article. The various cognitive requirements are to be found in all text functions and are balanced across each age appropriate test booklet (see Figure 1).

Figure 1: Text functions and cognitive requirements



4.3 Item formats

The majority of items are presented in a multiple-choice format. This type of item consists of a question related to a text with a choice of four possible answers of which one is correct. A further item format is decision-making items in which single statements have to be evaluated according to whether they are correct or incorrect in relation to the text. A third format is matching items in which a suitable heading has to be selected and assigned to each section of a text.[10] Second- and third-format items are summarized during the course of data analysis to produce items with partially correct solutions (for the scoring of the partial credit items; see Pohl & Carstensen, 2013, this issue). The different item formats should be applied across all text types, to as many cognitive requirements as possible, and across all age levels.

4.4 Test length and coverage of all features in the single tests

The reading competence test should take approximately 30 minutes to complete[11] per measurement time. In line with the framework, the final instruments for all age levels in the main national surveys each have five texts with the aforementioned five text functions and four to eight items for each text tapping local, deductive, and global comprehension (see, for concrete item examples in Grade 5 and 9, Gehrer, Zimmermann, Artelt, & Weinert, 2012). The three aforementioned cognitive requirements of the reading process in their eight different versions are balanced as far as possible in the test booklet for each starting cohort. All three item formats are applied in each test booklet.

By systematically considering the various text functions and operationalizing them in close-to-life and age-appropriate texts and text topics, and items tapping different cognitive requirements, it is possible to operationalize reading competence as a broadly based ability construct.

5. NEPS test development standards

The general challenge when implementing any measurement of competence lies in carefully transforming the theoretical foundations and framework model into empirically valid and reliable test instruments. The particular challenge when measuring reading competence longitudinally is to generate not only age-appropriate tests

10 Open unstructured formats are not used for both classificational and theoretical reasons. Measuring receptive language competencies (reading, listening) with productive procedures (e.g., writing or recounting a summary) is controversial in research, although short forms of open formats (e.g., composing titles) are in common use (e.g., PISA, EVAMAR).

11 Testing time also includes 2 minutes for self-ratings on the number of items solved (see Händel, Artelt, & Weinert, 2013, this issue, p. 170).

but also to test consistently and coherently across the lifespan. NEPS meets the demand for age-appropriate testing through an age-appropriate selection of stimulus texts and items; it meets the second demand for consistent and coherent modeling across the lifespan by consistently implementing the longitudinally designed framework.

To ensure that the text comprehension requirements correspond to the ability spectrum of the given age group, tests for all age groups are initially subjected to a careful preselection of test materials based on expert appraisals, difficulty indices, and readability indices. This should ensure that the stimulus texts are age-appropriate in terms of their length or shortness, style, syntax, vocabulary, and topics. This is followed by a multistage development and pretest process in which the items for each stimulus text are optimized successively in terms of their validity and model fit. One goal is for the items to provide an adequate spectrum of difficulty within each age group. How far this has been achieved for the first reading competence tests is one of the topics addressed in more detail in the empirical part of this article.

5.1 Empirical pilot studies to develop the test

Before they are applied in the field for the main survey, the instruments developed in NEPS have to go through several phases of cognitive interviews (see, for the method[12], Prüfer & Rexroth, 2005), smaller and larger preliminary studies, and large pilots (feasibility studies). Basically, the pool of texts and accompanying items developed for each starting cohort is four times as large as the final selection. Each development pool contains at least 20 texts and at least four examples of each text type – literary, commenting, advertising, instruction, and information. The complete pool of test material is piloted on the target population. One-half of each set of test materials is also tested on the next younger age cohort in NEPS[13] (e.g., items for 9th grade on a sample in 7th grade); the other half is also tested on the next higher age cohort (e.g., 12th grade). The samples for the pilots are representative samples from four German federal states (see, for a more detailed description of sampling procedures, Aßmann et al., 2011). After the pretests, suitable items are selected for each starting cohort in a two- to three-stage procedure. Items with less favorable parameters are optimized, and some new items are developed. Single units (i.e., texts and their attendant items) that prove to fit another age cohort better than that originally intended are reallocated. The resulting test materials are then given to the corresponding cohorts in pilot studies in four feder-

12 In the first phase of test development, cognitive interviews are suitable tools for obtaining early indications of problems that may arise (e.g., a question is not formulated clearly, certain words are not understood, text is found to be strenuous or boring, etc.).

13 In the school-age stages, reading competence is being measured every 2 years. Hence, in secondary school, it is being measured in the 5th, 7th, 9th, and, for some students, also in the 11th grade.

al states. The resulting data are, and will be, used to select the best items or units on the basis of content-related and statistical criteria. The statistical criteria[14] include fitting the items to the underlying unidimensional Rasch model, the absence of outliers for differential item functions (DIF) for gender and type of school, as well as item coverage of a broad spectrum of difficulty for one age group. The main emphasis is to cover all the criteria in the framework in a balanced way.

6. Study of the dimensionality of the reading competence test for adults

6.1 Research question

The main purpose of the features included in the framework for assessing reading competence is to cover the entire breadth of uses of reading and cognitive requirements. Due, among others, to the time constraints on testing within NEPS, the goal of test construction is to tap an intrinsically relatively homogeneous construct of reading competence in participants after psychometrically confirming the unidimensionality of the assessed construct. Items are also selected with reference to a unidimensional Rasch model. The assumption of unidimensionality is tested in this empirical section. Analyses of the dimensionality of reading comprehension tests are applied to data from a development study to ascertain whether the assumed unidimensionality can be confirmed, or whether and to what degree the text functions and cognitive requirements tap empirically distinguishable dimensions. This analysis is based on data from a pilot study with adults. It also tries to test the appropriateness of the framework using the text functions and cognitive requirements as structural features to assess reading competence: By analyzing the difficulty distribution of the items tapping the various text functions and cognitive requirements, we can analyze indirectly whether the items tapping the various levels of these features are suitable for assessing differences in ability between persons in this adult sample.[15] This should test empirically whether the test instrument resulting from the development study is appropriate for the main survey.

6.2 Sample

Participants were 447 adults (258 women and 186 men, 3 participants did not report gender). Our plan had been to stratify the sample to produce equal distributions of the variables age (birth cohorts Group 1: 1975–1989; Group 2: 1960–1974;

14 We thank Steffi Pohl for carrying out the basic scaling of the competence data for reading literacy (see, for a description of the respective analyses, Pohl & Carstensen, 2012).

15 This can already be assumed for 15-year-olds (PISA; see OECD, 2002).

Group 3: 1943–1959) and education (low, intermediate, high level of education[16]). However, we managed to only approach an equal distribution in the final sample: The youngest Group 1 contained 135 participants (30.2%); the intermediate Group 2, contained 146 (32.7%), and the oldest Group 3 contained 151 (33.8%) over 51-year-olds (18 participants did not report their age). The mean age was 43.3 years (*SD* = 2.4, range: 21.0–66.3 years). The lowest education group contained 107 persons (23.9%); 2 of these persons had left school without graduating, 2 had graduated from a special needs school [*Förderschulabschluss*], 58 had left school with basic school-leaving certificates [*einfacher Hauptschulabschluss*], and 45 with more qualified basic school-leaving certificates [*qualifizierender Hauptschulabschluss*]. The intermediate education group contained 159 persons (35.6%) with an intermediate school-leaving certificate [*Realschulabschluss*]. Group 3 with the highest education level contained 179 persons (39.8%), of whom 43 (9.6%) had a university of applied science entry certificate [*Fachhochschulreife*] and 135 (30.2%) had a university entrance certificate [*Abitur*]. Three participants gave no reports on their education.

6.3 Study implementation

The study was carried out by the *infas* survey institute. Participants were tested individually under standardized conditions in their own private homes. After completing a short questionnaire tapping the socio demographic origin, migration background and reading related behavior, they worked on four 30-minutes blocks, each containing five units consisting of one text and the accompanying items. They were paid € 15 (about $ 20) for their participation.

6.4 Instruments

A total of 26 units from the reading competence test were administered in the pilot study. By applying a multimatrix design, each participant received only 20 of the 26 units. The 26 units were distributed across 10 test booklets in which the sequence of units was also varied to balance out any effects of fatigue across all participants and tests.

The various text functions had an unequal distribution across the 26 units: There were 10 information texts, 5 commenting texts, 5 literary texts, 5 instruction texts, and 5 advertising texts. Because the greater number of information texts made the items in this category more heterogeneous, 5 of the 10 information texts were chosen at random for further analysis.

16 Level of education was determined by combining indicators on school education and the highest vocational training qualification.

We then performed an item selection. Exclusion criteria were a discriminative power of less than 0.20, an item misfit (MNSQ) greater than 1.25, and irregularities on the level of distractors (e.g., positive relation to the total test outcome). In addition, deviations in the item-characteristic curve (ICC) implied by the model from the data-based ICC were assessed qualitatively by the research team, which led to further items being dropped. A further criterion was the differential item functions (DIF) for gender and type of school. Items with extreme DIF values were dropped, whereas certain items with intermediate DIF values were retained when they were relevant for the content of a unit. This left 152 of the original 219 items. However, this set of items showed an overrepresentation of conclusion-drawing items, so a further randomly chosen 43 items were dropped from the dimension testing. Hence, the analyses presented here were based on 109 items that were roughly balanced in terms of the number of text types and the cognitive requirements (see Table 1).

Table 1: Distribution of text types and cognitive requirements in the 109-item analysis pool for testing multidimensionality

Text type/Function	Cognitive requirement			
	Finding information in text	Drawing text-related conclusions	Reflecting and assessing	∑ Items
Information	8	8	5	21
Commenting	6	6	9	21
Literary	3	9	8	20
Advertising	6	9	7	22
Instruction	10	6	9	25
∑ Items	33	38	38	109

6.5 Analysis strategy

The subsequent data analysis was performed with ACER Conquest 2.0 software (Wu, Adams, Wilson, & Haldane, 2008). The reading items were Rasch scaled (see Pohl & Carstensen, 2013, this issue). Complex items (decision-making and matching items) that were not answered completely correctly received partial credit scores. Three different models were computed with the item pool: a unidimensional model, a three-dimensional model (taking the three different cognitive requirements of the items into account), and a five-dimensional model (taking the five text types into account). We used two criteria from information theory (see Rost, 2004) as indicators for the comparative testing of the model fit: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). A further criterion was the significance of the deviance change.

7. Results on the dimensionality and difficulty of the adult reading competence test

7.1 Model fit

The results on whether the conceptual unidimensionality of reading competence assessment or the theoretical facets of the framework can be found in the empirical data structure present a differentiated picture (see Table 2). The dimension analyses showed that the text functions in particular showed clear empirical differences in the features, whereas the cognitive requirements are not verifiable as own empirical dimensions.

Table 2: Model comparison in terms of multidimensionality (N = 447, 109 Items)

Model	AIC	BIC	Deviance	Model parameters	Test versus unidimensional model
Unidimensional	35,306.33	35,970.94	34,982.33	162	–
Three-dimensional (cognitive requirements)	35,302.8	35,987.93	34,968.80	167	$\Delta\chi^2(5) = 13.53$, $p = .02$
Five-dimensional (functions of text)	35,166.03	35,888.08	34,814.03	176	$\Delta\chi^2(14) = 168.30$, $p = .00$

Note. AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion.

7.2 Cognitive requirements

The dimension analysis of the cognitive requirements showed that the three-dimensional model had a better numerical fit than the unidimensional model according to the AIC criterion and deviance change which attained statistical significance $\chi^2(5, N = 447) = 13.53$, $p = .02$ (see Table 2). However, the sample-weighted and therefore stricter BIC criterion had a higher value for the three-dimensional model (35,987.93) than for the unidimensional model (35,970.94), indicating a better fit with the data for the unidimensional model. Taken together, when stricter criteria were applied, the expected unidimensional construct of reading competence had a better fit than the three-dimensional model based on cognitive requirements.

All intercorrelations between the three dimensions of cognitive requirements adjusted for measurement error (see Table 3) were conspicuously high (r = .94, .96, and .99). The highest intercorrelation at r = .99 was between "reflecting and assessing" and the cognitive requirement "drawing text-related conclusions (during local and global coherence formation)". These two dimensions seem to form almost identical subfacets of reading competence in the NEPS test for adults.

Table 3: Intercorrelations between dimensions of cognitive requirements (N = 447, 109 Items)

	Finding information in text	Drawing text-related conclusions	Reflecting and assessing
Finding information in text	–		
Drawing text-related conclusions	.96	–	
Reflecting and assessing	.94	.99	–

Note. Intercorrelations are to be seen as correlations between the latent dimensions (as estimated in ACER Conquest 2.0).

7.3 Text functions and text types

The results of the dimension analyses showed that the five-dimensional model differentiating according to the text functions fitted the empirical data from the reading competence test better than the unidimensional model. The criteria were quite unequivocal: The deviance, the AIC, and sample-weighted BIC had the lowest values in the five-dimensional model compared to the unidimensional model. The deviance change criterion also attained statistical significance, $\chi^2(14, N = 447) = 168.30$, $p = .00$. Hence, the five-dimensional model based on text types was actually the best model. However, the intercorrelations corrected for measurement error between the five text functions were also very high (r = .78 to .91), although lower than those for cognitive requirements (see Table 4). Yet, there were some conspicuous findings here as well: The information texts correlated particularly strongly with the advertising (r = .91), the instruction (r = .89), and commenting texts (r = .87). There were also strong correlations of the instructions with advertising texts (r = .87) – a group of simple information texts containing appeals, calls, and announcements in which it is not necessary to follow complicated lines of argumentation. The literary texts correlated less strongly with the other texts, which is not surprising in light of the specific features of this category. With intercorrelations of r = .78 with information texts, r = .80 with instruction texts, r = .82 with commenting texts, and r = .83 with advertising texts, the literary texts showed a comparatively intermediate intercorrelation with the other text types.

Table 4: Intercorrelations between dimensions of text functions/text types (N = 447, 109 Items)

Text Functions/ Types	Information	Commenting	Literary	Instruction	Advertising
Information	–				
Commenting	.87	–			
Literary	.78	.82	–		
Instruction	.89	.85	.80	–	
Advertising	.91	.84	.83	.87	–

Note. Intercorrelations are to be seen as correlations between the latent dimensions (as estimated in ACER Conquest 2.0).

Apart from the intercorrelations, the analysis pool revealed that the most variance between readers came from the literary texts (σ^2 = 1.81). Hence, the differences between good and poor readers were markedly stronger here than in information (σ^2 = 1.5), advertising (σ^2 = 1.41), instruction (σ^2 = 1.63), or commenting texts (σ^2 = 1.41).

7.4 Selection of items for the main study and analysis of distribution of difficulty for the text functions and cognitive requirements

The 26 items selected from the preliminary study pool for the main study address five texts spread across all text functions and containing all three cognitive requirements. Basically, results showed that the selected items differentiated particularly well in the lower ability range, but not in the higher. Therefore, six further difficult items were modified or specially constructed for the main survey. However, because these were piloted after the present preliminary study pool, they cannot yet be presented with the same psychometrics.

An inspection of the distribution of item difficulties (ranging in total from -2.93 to 0.35 on the logit scale) as a function of text type revealed that the items in each text type – with the exception of commenting texts in which the items were closer together – had a difficulty spectrum ranging from 1 to 2.4 units on the logit scale. The difficulty spectrum covered by the items in each of the three cognitive requirements was higher than 1.8 units on the logit scale. For the requirement to find single pieces of information in text, the items even ranged over the entire difficulty spectrum of almost three logits for the items analyzed in this study.

8. Discussion

The specially developed NEPS reading competence test presented in this article is based on a framework oriented toward the standards of international LSAs of reading competence, and permit a longitudinal modeling of the development of reading competence over the lifespan. The NEPS framework for measuring reading competence unites the features of the cognitive requirements of the reading process with the conditions that different text functions and text types impose on readers. The tests hold both features constant across the lifespan. The presentation of the process of developing test instruments based on the NEPS framework showed how the quality of the future test instrument could be ensured by constructing a preliminary pool that was four to six times larger than needed, applying several phases of optimization and selection, and performing tests on several age cohorts. A randomly selected analysis pool[17] (109 items) from the pilot study for adults (N = 447) was used to examine whether the two features within the framework would also prove to be empirically distinguishable subdimensions of reading competence, or whether the expected unidimensionality of the assessment of reading competence could be confirmed. Results showed that it was particularly the cognitive requirements that had exceptionally high intercorrelations (r = .94, .96, and .99). This corresponded to the fact that a model comparison based on the stricter BIC criterion indicated that reading competence, as measured following the NEPS framework, is as hypothesized, a unidimensional rather than a three-dimensional construct and therefore the cognitive requirements are not verifiable as own empirical dimensions.

It seems worthwhile to compare the relatively high intercorrelations between the cognitive requirements with the relations between cognitive requirements found for the PISA reading literacy test (15-year-olds). Given that the intercorrelations of cognitive requirements in the PISA study were reported using uncorrected intercorrelations based on WLE (weighted likelihood) estimates, it is necessary to compare them to intercorrelations of the cognitive requirements in NEPS based on the same algorithm. The resulting uncorrected intercorrelations based on WLE estimators for the NEPS data for adults were markedly lower (r = .72, .77, and .71) and corresponded roughly with the uncorrected correlations in PISA (r = .74, .71, and .64). The comparison confirmed, among others, the smallest correlation found in the NEPS data between the subdimensions “finding information” and “reflecting and assessing”. However, whereas in PISA, the closest relation was found between the subdimensions “finding information” and “text-related interpreting” (Artelt & Schlagmüller, 2004), NEPS finds the closest relation between the cognitive requirements “drawing text-related conclusions” and “reflecting and assessing”.

17 Because only five texts each accompanied by 4–6 items can be applied in the main surveys due to time restrictions, the multidimensionality testing was performed with a larger pool from the development study. The analysis pool was refined by removing items that failed to reach test validity criteria and reducing the overhang of information texts. The resulting dimension analyses could be performed on the basis of 109 aggregated items.

When interpreting these different subfindings, it is necessary to recall that although NEPS applies a similar framework for cognitive requirements to that in PISA, it is not identical. For example, in contrast to PISA, the NEPS framework places more emphasis on the local and global coherence formation in the subdimension "drawing text-related conclusions", and less emphasis on integrating background knowledge in the different levels of "reflecting and assessing". Naturally, these theoretical differences in the framework influence the test construction and hence the results of both studies. In addition, the concentration on continuous texts in NEPS versus the addition of discontinuous texts in PISA also has an appreciable effect on the cognitive requirements.

It is still necessary to test whether the markedly high intercorrelation of r = .77/.99 between "reflecting and assessing" and "drawing text-related conclusions", is also due to occasional slight overlaps in the requirements of the subtypes of items. For example, the drawing text-related conclusions task of "being able to understand the important ideas in a text on the basis of comprehending more complex relevant text sections" (Type 2.3 items) and the reflecting and assessing task of "comprehending the central idea, event, or message of a text" (Type 3.1) also seem to have very similar contents.

The test of multidimensionality based on text type features showed major differences between the text functions. The fact, that the model based on the different text types showed a better model fit than the hypothesized unidimensional model of the framework, could provide an indication in the direction – as supposed from academics in the domain of languages and educational didactic –, that the different features of text types exert an influence on their reception and connected reading achievement.

The strong relation between information texts and all other text types apart from literary texts indicates that – greatly simplified – all text types used apart from literary texts can be conceived more generally as information or instruction texts; that is, the typical requirements they place on reading competence have much in common. These result concerning the difference between literary versus information based subcompetencies was already presented by the detailed analysis of the PISA data (Artelt & Schlagmüller, 2004). New results in the domain of literary-aesthetic research confirm this view; Roick and colleagues showed that for adolescents literary versus information based subcompetencies can be distinguished empirically (Roick, Frederking, Henschel, & Meier, 2013). Though both studies operationalized literary reading above all of the three literary genres – the lyrics, the drama and the epos or narrative literature –, meanwhile the NEPS framework and the reading test instruments focus exclusively on the narrative literature, remarkably, our first results in analyzing an adult pilot study goes in the same direction.

Our analysis of the text types additionally confirmed the constructional assumption linked to the NEPS framework that a heterogeneous selection of continuous text types can provide a balanced assessment of reading competence in the sense of text comprehension. An inspection of the distribution of item difficulties as a func-

tion of the text type or the specific cognitive requirement reveals that each text type or requirement delivers a satisfactory to very good differentiation in this age range.

The presentation of the items intended for the main survey revealed that the selected items differentiated particularly well in the lower ability range but not in the upper range. This finding can and has to be qualified by stating that the sample in this preliminary study was a positive selection due to the failure to achieve an equal distribution of educational attainments. It can be assumed that this led to a tendency to overestimate the performance of adults in the population and underestimate the difficulty of the items. Nonetheless, additional difficult items were constructed to tap the upper ability range more broadly.

Finally, it should be pointed out that the analyses carried out here are still only provisional trends until the analyses of further age cohorts (Grades 5 and 9 and college students) deliver comparative confirmation of the results obtained for the adult items. Particularly for highly correlated dimensions, Harell (2009), for example, has pointed out that the use of information theory criteria does not always suffice to identify the right model.[18]

Moreover, the multidimensionality of the two category systems should not just be interpreted in a disassociated way. A promising next step in this regard would be to perform an in-depth multidimensionality analysis assuming within-item multidimensionality (see Hartig & Höhler, 2008).

Acknowledgments

This work was supported by third-party funds from the German Federal Ministry of Education and Research (BMBF). We thank Franziska Fellenberg, NEPS, University of Bamberg, for her assistance in the conception and development of the reading test for the starting cohorts, and Jonathan Harrow for translating this article into English. The original, unpublished German-language version is available from the first author on request.

References

Abraham, U. (2003). Lese- und Schreibstrategien im themazentrierten Deutschunterricht. Zu einer Didaktik selbstgesteuerten und zielbewussten Umgangs mit Texten. In U. Abraham, A. Bremerich-Vos, V. Frederking, & P. Wieler (Eds.), *Deutschdidaktik und Deutschunterricht nach PISA* (pp. 204–219). Breisgau, Germany: Filibach.

Adamzik, K. (2010). Sprachwissenschaftliche Gattungsforschung. In R. Zymner (Ed.), *Handbuch Gattungstheorie* (pp. 295–298). Stuttgart, Germany: Metzler.

18 However, the simulation studies in which he was able to show this applied relatively short instruments (20 items and 40 items). His results indicate that a longer instrument is needed for highly correlated dimensions – longer like the ones used in the present study.

Artelt, C., McElvany, N., Christmann, N., Richter, T., Groeben, N., Köster, J., Schneider, W., Stanat, P., Ostermeier, C., Schiefele, U., Valtin, R., Ring, K., & Saalbach, H. (2005). *Expertise – Förderung von Lesekompetenz* (Bildungsreform No. 17). Berlin/Bonn, Germany: Bildungsministerium für Bildung und Forschung.

Artelt, C., Schiefele, U., Schneider, W., & Stanat, P. (2002). Leseleistungen deutscher Schülerinnen und Schüler im internationalen Vergleich (PISA): Ergebnisse und Erklärungsansätze. *Zeitschrift für Erziehungswissenschaft, 5*(1), 6–27.

Artelt, C., & Schlagmüller, M. (2004). Der Umgang mit literarischen Texten als Teilkompetenz im Lesen? Dimensionsanalysen und Ländervergleiche. In U. Schiefele, C. Artelt, W. Scheider, & P. Stanat (Eds.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (pp. 169–190). Wiesbaden, Germany: VS.

Artelt, C., Stanat, P., Schneider, W., & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, J. Tillmann, & M. Weiß (Eds.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 69–137). Opladen, Germany: Leske + Budrich.

Artelt, C., Stanat, P., Schneider, W., Schiefele, U., & Lehmann, R. H. (2004). Die PISA-Studie zur Lesekompetenz. Überblick und weiterführende Analysen. In U. Schiefele, C. Artelt, W. Scheider, & P. Stanat (Eds.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (pp. 139–168). Wiesbaden, Germany: VS.

Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., Rohwer, G., Rässler, S., & Blossfeld, H.-P. (2011). Sampling designs of the National Educational Panel Study: Challenges and solutions. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Education as a lifelong process. The German National Educational Panel Study (NEPS). Zeitschrift für Erziehungswissenschaft [Special Issue 14]* (pp. 51–65). Wiesbaden, Germany: VS.

Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Schümer, G., Stanat, P., Tillmann, J., & Weiß, M. (Eds.). (2002). *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich: Zusammenfassung zentraler Befunde*. Opladen, Germany: Leske + Budrich.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, J., & Weiß, M. (Eds.). (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen, Germany: Leske + Budrich.

Baumert, J., Stanat, P., & Demmrich, A. (2001). Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, J. Tillmann, & M. Weiß (Eds.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 15–68). Opladen, Germany: Leske + Budrich.

Beck, B., & Klieme, E. (Eds.). (2003). *DESI – Eine Längsschnittstudie zur Untersuchung des Sprachunterrichts in deutschen Schulen. Empirische Pädagogik*. Weinheim, Germany: Beltz.

Beck, B., & Klieme, E. (2007a). Einleitung. In B. Beck & E. Klieme (Eds.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (pp. 1–8). Weinheim, Germany: Beltz.

Beck, B., & Klieme, E. (Eds.). (2007b). *Sprachliche Kompetenzen: Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)*. Weinheim, Germany: Beltz.

Berendes, K., Weinert, S., Zimmermann, S., & Artelt, C. (2013) Assessing language indicators across the lifespan within the German National Educational Panel Study (NEPS). *Journal for Educational Research Online, 5*(2), 15–49.

Blossfeld, H.-P., Roßbach, H.-G., & Maurice, J. von (Eds.). (2011). *Education as a lifelong process: The German National Educational Panel Study (NEPS). Zeitschrift für Erziehungswissenschaft [Special Issue 14]*. Wiesbaden, Germany: VS.

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., & Walther, G. (Eds.). (2004). *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster, Germany: Waxmann.

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G., & Valtin, R. (Eds.). (2003). *Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster, Germany: Waxmann.

Bremerich-Vos, A., & Grotjahn, R. (2007). Lesekompetenz und Sprachbewusstheit: Anmerkungen zu zwei aktuellen Debatten. In B. Beck & E. Klieme (Eds.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (pp. 158–177). Weinheim, Germany: Beltz.

Brinker, K. (1983). Textfunktionen. Ansätze zu ihrer Beschreibung. *Zeitschrift für germanistische Linguistik (ZGL), 11*, 127–148.

Brinker, K. (1985). *Linguistische Textanalyse: Eine Einführung in Grundbegriffe und Methoden*. Berlin, Germany: Filibach.

Brinker, K. (2010). *Linguistische Textanalyse: Eine Einführung in Grundbegriffe und Methoden* (7th ed.). Berlin, Germany: ESV.

Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena, Germany: Fischer.

Christmann, U., & Groeben, N. (1999). Psychologie des Lesens. In B. Franzmann, K. Hasemann, D. Löffler, & E. Schön (Eds.), *Handbuch Lesen. Im Auftrag der Stiftung Lesen und der Deutschen Literaturkonferenz* (pp. 145–223). München, Germany: Saur.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment* [Adobe Reader version]. Retrieved from http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

DESI Konsortium. (2006). *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Zentrale Befunde der Studie Deutsch-Englisch-Schülerleistungen-International (DESI)*. Frankfurt a. M., Germany: Deutsches Institut für Internationale Pädagogische Forschung (DIPF).

Eberle, F., Gehrer, K., Jaggi, B., Kottonau, J., Oepke, M., & Pflüger, M. (2008). *Evaluation der Maturitätsreform 1995 (EVAMAR): Phase II*. Bern, Switzerland: Staatsekretariat für Bildung und Forschung.

Garbe, C., Holle, K., & Jesch, T. (2009). *Texte lesen. Lesekompetenz – Textverstehen – Lesedidaktik – Lesesozialisation*. Paderborn, Germany: Schöningh.

Gehrer, K., & Artelt, C. (2013). Literalität und Bildungslaufbahn: Das Bildungspanel NEPS. In A. Bertschi-Kaufmann & C. Rosebrock (Eds.), *Literalität erfassen: bildungspolitisch, kulturell, individuell*. Weinheim, Germany: Juventa.

Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2012). *The assessment of reading competence (including sample items for grade 5 and 9). Scientific Use File 2012, Version 1.0.0*. Bamberg, Germany: University of Bamberg, National Educational Panel Study. Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/1-0-0/com_re_2012_en.pdf

Göttsche, D. (2010). Prosa als Bestimmungskriterium. In R. Zymner (Ed.), *Handbuch Gattungstheorie* (pp. 38–39). Stuttgart, Germany: Metzler.

Groeben, N., & Hurrelmann, B. (2004). *Lesesozialisation in der Mediengesellschaft: Ein Forschungsüberblick*. Weinheim, Germany: Juventa.

Grotlüschen, A., Kretschmann, R., Quante-Brandt, E., & Wolf, K. D. (Eds.). (2011). *Literalitätsentwicklung von Arbeitskräften* (Alphabetisierung und Grundbildung, Vol. 6). Münster, Germany: Waxmann.

Grotlüschen, A., & Riekmann, W. (2011). *leo. – Level-One Studie: Literalität von Erwachsenen auf den unteren Kompetenzniveaus (Presseheft)*. Hamburg, Germany: Universität Hamburg.

Händel, M., Artelt, C., & Weinert, S. (2013). Assessing metacognitive knowledge: Development and evaluation of a test instrument. *Journal for Educational Research Online, 5*(2), 162–188.

Harell, L. M. (2009). *Accuracy of global fit indices as indicators of multidimensionality in multidimensional Rasch analysis* (Doctoral dissertation). Blacksburg, VA: Virginia Polytechnic Institute and State University.

Hartig, J., & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within-item and between-item multidimensionality. *Zeitschrift für Psychologie, 216*(2), 89–101.

Hess-Lüttich, E. (Ed.). (1996). *Textstrukturen im Medienwandel*. Frankfurt a. M., Germany: Lang.

Kintsch, W. (1982). *Gedächtnis und Kognition*. Berlin, Germany: Springer (Original work published 1970).

Kintsch, W. (1994). Kognitionspsychologische Modelle des Textverstehens: Literarische Texte. In K. Reusser & M. Reusser-Weyeneth (Eds.), *Verstehen. Psychologischer Prozess und didaktische Aufgabe* (pp. 39–54). Bern, Switzerland: Hans Huber.

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.

Kirsch, I. S. (1995). Literacy performance on three scales: Definitions and results. In OECD & Statistics Canada (Eds.), *Literacy, economy and society. Results of the first International Adult Literacy Survey* (pp. 27–53). Paris, France: OECD.

Kirsch, I. S. (2001). *The International Adult Literacy Survey (IALS). Understanding what was measured*. Princeton, NJ: Research Publications Office.

Kirsch, I. S., Jungeblut, A., & Mosenthal, P. B. (1998). The measurement of adult literacy. In T. S. Murray, I. S. Kirsch, & L. Jenkins (Eds.), *Adult literacy in OECD countries: Technical report on the first International Adult Literacy Survey*. Washington, DC: US Department of Education, National Center for Education Statistics.

Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (Eds.). (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster, Germany: Waxmann.

Köster, J. (2003). Die Bedeutung des Vorwissens für die Lesekompetenz. In U. Abraham, A. Bremerich-Vos, V. Frederking, & P. Wieler (Eds.), *Deutschdidaktik und Deutschunterricht nach PISA* (pp. 90–104). Breisgau, Germany: Filibach.

Mosenthal, P. B. (1996). Understanding the strategies of document literacy and their conditions of use. *Journal of Educational Psychology, 88*(2), 314–332.

Mosenthal, P. B., & Kirsch, I. S. (1991). Toward an explanatory model of document literacy. *Discourse Processes, 14*(2), 147–180.

Mosenthal, P. B., & Kirsch, I. S. (1998). A new measure for assessing document complexity: The PMOSE/IKIRSCH document readability formula. *Journal of Adolescent and Adult Literacy, 41*(8), 638–657.

Mullis, I. V. S., Martin, M. O., González, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary school in 35 countries*. Chestnut Hill, MA: Boston College.

Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Reading competence von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider, & P. Stanat (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt* (pp. 23–72). Münster, Germany: Waxmann.

Nold, G., & Willenberg, H. (2007). Lesefähigkeit. In B. Beck & E. Klieme (Eds.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (pp. 23–41). Weinheim, Germany: Beltz.

OECD – Organisation for Economic Co-Operation and Development. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris, France: OECD.

OECD – Organisation for Economic Co-Operation and Development. (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. Retrieved from http://www.oecd.org/dataoecd/44/53/33691596.pdf

OECD – Organisation for Economic Co-Operation and Development. (2002). *PISA 2000 technical report*. Paris, France: OECD.

OECD – Organisation for Economic Co-Operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris, France: OECD. Retrieved from www.oecd.org/dataoecd/16/52/35888923.pdf

OECD – Organisation for Economic Co-Operation and Development. (2007). *Kurzzusammenfassung PISA 2006*. Retrieved from www.oecd.org/data oecd/18/35/39715718.pdf

OECD – Organisation for Economic Co-Operation and Development. (2009). *PISA 2009 assessment framework – Key competencies in reading, mathematics, and science*. Paris, France: OECD.

OECD – Organisation for Economic Co-Operation and Development. (2010). *PISA 2009 results: Executive summary [PISA 2009 Ergebnisse: Zusammenfassung]*. Retrieved from www.oecd.org/dataoecd/34/19/46619755.pdf

OECD – Organisation for Economic Co-Operation and Development, & Statistics Canada. (1995). *Literacy, economy and society: Results of the first International Adult Literacy Survey*. Paris, France: OECD.

Pohl, S., & Carstensen, C. (2012). *NEPS technical report – Scaling the data of the competence tests (NEPS Working Paper No. 14)*. Bamberg, Germany: Otto-Friedrich-Universität, Nationales Bildungspanel.

Pohl, S., & Carstensen, C. (2013). Scaling of competence tests in the National Educational Panel Study – Many questions, some answers, and further challenges. *Journal for Educational Research Online, 5*(2), 189–216.

Prüfer, P., & Rexroth, M. (2005). Kognitive Interviews. *ZUMA How-to-Reihe, 15*, 1–21.

Quathamer, D. (1998). *Kohärenzbildung beim Lesen von Texten – Nutzung und Funktion von Überblicksdiagrammen* (Doctoral dissertation). Duisburg, Germany: Gerhard-Mercator-Universität.

Richter, T., & Christmann, U. (2002). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Eds.), *Lesekompetenz: Bedingungen, Dimensionen, Funktionen* (pp. 25–58). Weinheim, Germany: Juventa.

Roick, T., Frederking, V., Henschel, S., & Meier, C. (2013). Literarische Textverstehenskompetenz bei Schülerinnen und Schülern unterschiedlicher Schulformen. In A. Bertschi-Kaufmann & C. Rosebrock (Eds.), *Literalität erfassen: bildungspolitisch, kulturell, individuell* (pp. 69–85). Weinheim, Germany: Juventa.

Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2nd ed.). Bern, Switzerland: Hans Huber.

Schaffner, E., Schiefele, U., Drechsel, B., & Artelt, C. (2004). Lesekompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost, & U. Schiefele (Eds.), *PISA 2003 – Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (pp. 93–110). Münster, Germany: Waxmann.

Scheffel, M. (2010a). Faktualität/Fiktionalität als Bestimmungskriterium. In R. Zymner (Ed.), *Handbuch Gattungstheorie* (pp. 29–31). Stuttgart, Germany: Metzler.

Scheffel, M. (2010b). Theorien des Narrativen. In R. Zymner (Ed.), *Handbuch Gattungstheorie* (pp. 328–331). Stuttgart, Germany: Metzler.

Schnotz, W. (1988). Textverstehen als Aufbau mentaler Modelle. In H. Mandl & H. Spada (Eds.), *Wissenspsychologie* (pp. 299–330). München, Germany: PVU.

Schnotz, W., & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Text- und Bildverstehen. *Zeitschrift für Experimentelle Psychologie, 46*(3), 217–236.

Schnotz, W., & Dutke, S. (2004). Kognitionspsychologische Grundlagen der Lesekompetenz: Mehrebenenverarbeitung anhand multipler Informationsquellen. In U. Schiefele, C. Artelt, W. Scheider, & P. Stanat (Eds.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (pp. 61–99). Wiesbaden, Germany: VS.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C. (2011). Development of competencies across the lifespan. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Education as a lifelong process. The German National Educational Panel Study (NEPS). Zeitschrift für Erziehungswissenschaft [Special Issue 14]* (pp. 67–86). Wiesbaden, Germany: VS.

Werlich, E. (1975). *Typologie der Texte: Entwurf eines textlinguistischen Modells zur Grundlegung einer Textgrammatik*. Heidelberg, Germany: Quelle & Meyer.

Werlich, E. (1976). *A text grammar of English*. Heidelberg, Germany: Quelle & Meyer.

Wieler, P. (2003). Varianten des Literacy-Konzepts und ihre Bedeutung für die Deutschdidaktik. In U. Abraham, A. Bremerich-Vos, V. Frederking, & P. Wieler (Eds.), *Deutschdidaktik und Deutschunterricht nach PISA* (pp. 47–68). Breisgau, Germany: Filibach.

Willenberg, H. (2007). Lesen. In B. Beck & E. Klieme (Eds.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (pp. 107–117). Weinheim, Germany: Beltz.

Wu, M. L., Adams, R. J., Wilson, M., & Haldane, S. (2008). *ACER Conquest 2.0: Generalized item response modeling software* [Computer program]. Hawthorn, Australia: ACER.

Zymner, R. (Ed.). (2010). *Handbuch Gattungstheorie*. Stuttgart, Germany: Metzler.