

# Assessment of metacognitive knowledge in students with special educational needs

Marion Händel, Kathrin Lockl, Jana Heydrich,  
Sabine Weinert & Cordula Artelt

**Abstract:** This study investigated whether and, if so, how metacognitive knowledge can be assessed validly in students with special educational needs in a large-scale assessment like the German National Educational Panel Study. In total, 804 sixth-grade students including both regular school students attending the lowest track of secondary education (Hauptschule) and students with special educational needs in learning participated in the study. A scenario-based test of metacognitive knowledge focusing primarily on different aspects of strategy knowledge was implemented. In order to investigate optimal testing conditions, two conditions that varied in terms of administration mode were compared: *autonomous reading* as in regular test settings and a *read-aloud* condition. Reading speed and reasoning abilities were assessed as control variables. As expected, regular school students outperformed students with special educational needs in the metacognitive knowledge test. In addition, higher correlations between metacognitive knowledge and reading speed emerged in the autonomous reading condition compared to the read-aloud condition. Contrary to our expectations, a differential boost due to the testing accommodation of reading aloud was, however, only observed in regular students but not in students with special educational needs. The results are discussed with regard to educational and assessment-relevant approaches.

**Keywords:** Metacognitive knowledge, Special educational needs, Testing accommodations, Large-scale study

The importance of metacognition is shown by its relationship to measures of ability and achievement (van Kraayenoord and Schneider 1999; Pressley et al. 1989). Students who self-regulate their learning or have more sophisticated metacognitive knowledge perform better at school than their counterparts with poorer self-regulatory learning behavior or less metacognitive knowledge. Moreover, deficits in metacognitive knowledge and metacognitive processes are supposed to play an important role in the emergence of learning difficulties in students with special educational needs (Hannah and Shore 1995; Pintrich et al. 1994). As a consequence, the valid assessment of metacognitive knowledge is of particular interest for both research and education. In the following, the focus is on metacognitive knowledge and its assessment, especially in students with special educational needs in learning (SEN-L). We report on a feasibility study of the German National Educational Panel Study (NEPS). The study investigated whether and, if so, how valid indicators of metacognitive knowledge can be obtained when students with SEN and regular school students are involved in a large-scale assessment.

## Students with special educational needs in learning (SEN-L)

In Germany, special schools are established for students with SEN. Within the group of students with SEN enrolled in special schools or school centers in Germany, students with SEN-L constitute the largest subsample (approximately 45 %) (Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany 2008). Students with SEN-L seem to have a lower achievement potential compared to students attending regular schools (Wocken and Gröhlich 2009). They are supposed to not being able to achieve the aims and contents of regular school (Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany 1999). Students attending special schools for SEN-L are more often male and mostly come from lower class families with many children (Bos et al. 2010; Statistisches Bundesamt 2012). They constitute a relatively heterogeneous group of students. One reason might be that there are no clear rules for the diagnostic process

in Germany (Bos et al. 2010). In addition, there is no international or German definition of SEN-L that is precise or consistent<sup>1</sup> (Bos et al. 2010; Hammill 1990; Heydrich et al. 2013; Twomey 2006). Explicitly excluded from the group of students with SEN-L in Germany and educated at different types of special schools are students with visual impairments, hearing disability/impairment, specific language/speech impairments, physical handicaps/disabilities, severe intellectual impairment/disability, emotional and behavioral difficulties, comprehensive SEN, and students with health impairment. It is widely assumed that students with SEN-L are especially lacking metacognitive competencies that are needed to regulate one's own learning and the acquisition of other competencies (Reid et al. 2004). Students with SEN-L seem to exhibit problems related to the regulation of learning, for example, within the spontaneous use of mnemonic skills or the flexible and reflective monitoring and regulation of one's own memory behavior (Brown 1978; Meltzer et al. 2004; Schröder 2000). Hannah and Shore (1995) summarized that "their general learning disabilities are due in part to deficiencies in metacognitive processes" (p. 96). In addition to procedural aspects of metacognition, metacognitive knowledge as the declarative aspect of metacognition is supposed to constitute a necessary prerequisite for strategic learning. Accordingly, students' failure to use and to generalize learned strategies might be seen as lacking metacognitive knowledge. Consistent with this assumption, results by Pintrich et al. (1994) revealed that at least some of the investigated students with learning disabilities were performing poorly in reading because they lacked metacognitive knowledge about reading strategies.

## Metacognitive knowledge and its assessment

Metacognitive knowledge refers to the knowledge about memory, comprehension, and learning processes that an individual can verbalize. Flavell (1979) defined metacognitive knowledge as knowledge about persons, tasks, and strategies. Metacognitive knowledge about strategies means knowledge about methods of how to learn effectively. It can further be subdivided into declarative (knowledge about the existence of certain strategies), procedural (knowledge about how a strategy works effectively), and conditional (knowledge about strategies that are useful for solving a certain task) strategy knowledge (Paris et al. 1983). The relevance of strategy knowledge has been both emphasized theoretically – for example, in the model of good information processing (Pressley et al. 1989) – and demonstrated empirically – for example, in the

child development research on meta-memory (see Schneider and Pressley 1997, for a review).

To assess metacognitive knowledge – that is, the knowledge about strategies, their use, and usefulness in certain situations – different procedures can be applied. Assessment methods that refer to interview situations (Kreutzer et al. 1975; Myers and Paris 1978) and think-aloud protocols (Swanson 1990) are expensive in implementation and analysis. In addition, they do not allow for a test administration in a group setting, which is essential for large-scale studies. In order to obtain valid indicators of metacognitive knowledge within group settings, tests concerned with children's knowledge about memory processes were developed (Belmont and Borkowski 1988; Hasselhorn 1994; Körkel 1987). These early test instruments, however, rely partially on fuzzy conceptualizations and exhibit rather unsatisfactory internal consistencies (Hasselhorn 1994; Kurtz et al. 1982). Initial paper-and-pencil multiple-choice instruments about conditional knowledge of reading strategies came from Jacobs and Paris (1987) and Schmitt (1990). These tests were further extended. Recently developed tests consist of several scenarios describing specific challenging situations; each scenario followed by a list of approaches of differing strategic quality that have to be rated according to their usefulness in the respective situation. Examples of such tests are metacognitive knowledge tests about reading strategies (Schlagmüller and Schneider 2007; see Artelt et al. 2009, for the items implemented in the PISA studies). Other test instruments refer to several content domains, such as the tests on metacognitive knowledge developed by Maag Merki et al. (2013) or Neuenhaus et al. (2011). These tests are suitable for a standardized group setting, exhibit a clear benchmark criterion, and focus primarily on different aspects of strategy knowledge; see also the NEPS test by Händel et al. (2013) that was used in this study.

In contrast to group-administered tests that students have to read and then complete by themselves, the assessment of metacognitive knowledge in students with SEN has mostly been conducted in rather uneconomic individual settings via interviews (Bosson et al. 2010). So far, little is known about metacognitive competencies of students with SEN-L from large-scale studies. If students with SEN-L were involved in large-scale studies (they have often been excluded in earlier times; see Elbaum et al. 2004; Johnson 2000; Ketterlin-Geller et al. 2007), it was often for reasons of representativeness (see the PISA studies; Hörmann 2007; von Stechow 2006). Several efforts were undertaken to give all individuals the opportunity to participate in large-scale assessments (especially in the

---

<sup>1</sup> As for the term "SEN-L", the term "learning disabilities; LD" is not clearly defined. For example, according to Fletcher (2012) "there is no consensus over which attributes are best for defining LD" (p. 20). Although the terms do not depict the same concepts (Schröder 2000) they describe overlapping populations of students. Accordingly, the investigated group of German students with SEN-L might include some but not all students who would be characterized as students with learning disabilities. For example, the concept of learning disabilities ("Lernschwierigkeiten") is broader and additionally includes students with rather high IQ but comparably low school achievement, so-called underachiever (cf. Schröder 2000). However, both terms refer to heterogeneous groups with multifaceted etiology.

United States of America, e.g., No Child Left behind Act, Americans with Disabilities Act, National Assessment of Educational Progress; see Pitoniak and Royer 2001). Nevertheless, Hollenbeck (2002) concluded that “the amount of experimental research conducted on accommodations is minimal” (p. 402). In addition, the focus was on cognitive rather than metacognitive competencies. To provide a valid assessment of metacognitive knowledge in students with SEN-L involved in large-scale assessments, it has to be considered that students with SEN-L often struggle with other, construct-irrelevant difficulties. These difficulties might pose a barrier and lead to unfair testing conditions. Therefore, to establish fair testing situations for all participating students, construct-irrelevant disabilities have to be taken into account.

## Test accommodation and validity aspects

Test scores should refer to those constructs that are intended to be assessed (Koretz and Barton 2003; McDonnell et al. 1997; Sireci et al. 2005). In order to minimize the influence of construct-irrelevant barriers for students with SEN and to increase test fairness, test accommodations<sup>2</sup> can be applied (Ketterlin-Geller et al. 2007; Pitoniak and Royer 2001; Sireci et al. 2005). By implementing accommodations the participation rates in nationwide and state assessments can be increased (NCLD – National Center for Learning Disabilities 2007). For instance, using written material with students with SEN might itself distort measurement and interpretations by negatively biased test scores due to construct-irrelevant reading disabilities (Koretz and Barton 2003). By way of example, students might not be able to complete the test in the time set. So far, accommodations have only been investigated for the assessment of cognitive competencies. Although some studies have implemented test accommodations such as reading-aloud for paper-and-pencil based tests of metacognitive knowledge (Swanson and Trahan 1996), they did not investigate how this impacted performance. Still, it is “unclear about what constitutes an appropriate and effective accommodation” (Sáez et al. 2013, p. 126).

Common accommodations<sup>3</sup> in test administration for students with SEN are shortened instruments completed in the same testing time, more processing time for the same amount of items, out-of-level-testing – that is, using items initially designed for younger students, or reading items and/or directions aloud (see Bolt and Ysseldyke 2007; Koretz and Barton 2003; Pitoniak and Royer 2001 for a discussion of

different test accommodations in students with SEN). The various accommodations result in different changes to test processing or the test instrument. Shortened instruments lead to design-related changes due to excluded items or item blocks. As time limits are essential for speed and power tests, a deviation from the test time would lead to construct-relevant changes. In addition, the accommodation of testing time might result in different effects depending on the individual speed of information processing and the ability to sustain attention over longer periods of time. Implementing out-of-level items allows for the comparison with younger students – as long as they are also part of the investigation – but may hamper the comparability with students of the same age/grade. Moreover, implementing different items leads to content-related changes. Accommodations that do not influence the design, the construct, or the content of the domain under investigation but are still supposed to facilitate test processing for students with SEN-L are accommodations that refer to the reading burden. Hence, written tests in group settings might lead to disadvantages for weak readers (cf. Kubinger 2009). Therefore, reading the tasks and items aloud to the children might be a reasonable accommodation.

Some accommodations may enable test processing for all kinds of students. When the special, construct-irrelevant disabilities of students with SEN are taken into account, however, test accommodations should result in a positive effect on students with SEN-L only (Fuchs and Fuchs 1999). That is, to provide a valid assessment (see Johnson 2000 for the discussion of several aspects that need consideration for validity), the test accommodations should have a *differential boost* for this group of students (Bolt and Ysseldyke 2007). Hence, it is important to consider the effects of test accommodations with respect to students with SEN-L as well as other groups of students. However, a differential boost might be produced by ceiling effects for regular students who do not need any accommodation at all (Laitusis 2010). Previous studies focusing on students with SEN-L in Germany used different kinds of test accommodations (Lehmann and Hoffmann 2009; Wocken 2000; Wocken and Gröhlich 2009). However, the studies did not allow for comparisons with other groups of students. Therefore, within a feasibility study of the NEPS, we investigated how reading items and instructions aloud would influence the performance on a metacognitive knowledge test taken by two groups of students: regular students attending the lowest track of secondary education in Germany<sup>4</sup> and students with SEN-L attending special schools.

---

<sup>2</sup> Note that some authors use accommodation and modification interchangeably whereas others differentiate between them: While accommodations are not meant to change the nature of the construct being measured, modifications result in a change in the test and equally affect all students taking it (Hollenbeck 2002).

<sup>3</sup> For an overview of further accommodations, also for students with other disabilities, we refer to Cormier et al. (2010) or Thurlow (2002).

<sup>4</sup> In most German Federal States school tracking is accomplished by way of ability tracking after Grade 4.

## The German National Educational Panel Study (NEPS)

The NEPS is a national, large-scale longitudinal study (comprising a multi-cohort sequence design) that investigates the development of competencies across the life span (Blossfeld et al. 2011; Blossfeld and von Maurice 2011). The study aims to provide high-quality, user-friendly data on competence development and educationally relevant processes for an international scientific community (Barkow et al. 2011). The competence domains under study are language competencies, mathematical competence, and scientific literacy, as well as meta-competencies, and social competencies (see Weinert et al. 2011 for more detailed information). Students attending regular schools as well as students attending special schools are surveyed as part of the NEPS (cf. Heydrich et al. 2013).

In order to make a substantial contribution to exploring students with SEN-L attending special schools, this group has been oversampled<sup>5</sup>. The NEPS is attempting to balance the requirements of students with SEN-L on the one hand and the methodological and statistical requirements of gathering valid and comparable large-scale data on the other hand. For the regular NEPS student cohorts, metacognitive knowledge is assessed using a scenario-based metacognitive knowledge test in Grades 1, 3, 6, and 9. A feasibility study in Grade 6 investigates whether and, if so, how metacognitive knowledge can be validly assessed in students with SEN-L in comparison to regular school students. As it is known that students with SEN-L, on average, show low reading competencies (Thurlow 2010; Thurlow et al. 2008; Ysseldyke et al. 1998), one key challenge for the group-administered assessment of metacognitive knowledge in students with SEN-L is that the test requires reading abilities.

### Aims of the study and research questions

Due to limited evidence about metacognitive knowledge of students with SEN-L the study pursued the following three aims and related research questions. (1) The study aimed to examine whether, and if so, how metacognitive knowledge can be validly assessed via a scenario-based test within the NEPS. The study assessed metacognitive knowledge scores in students

with SEN-L in a regular test condition. Furthermore, it was investigated whether a test accommodation impacts the metacognitive knowledge scores of students with SEN-L. (2) The study intended to test for potential differences between low-achieving students attending regular schools (Hauptschule, i.e., the lowest track of secondary education) and students attending special schools for students with SEN-L. In particular, does the implemented accommodation provide a differential boost to students with SEN-L compared to low-performing regular students? (3) Finally, the extent to which the measurement of metacognitive knowledge would depend on other skills such as reading speed was considered in both groups of students.

## Method

### Sample

In total, 804 sixth-grade students in two different school types from several Federal States in Germany participated in the feasibility study of the NEPS. Approximately half of them were students with SEN-L attending special schools,<sup>6</sup> whereas the other half of them were regular students from the lowest track of secondary education in Germany (“Hauptschule”), serving as a control group. Students of both school types are assumed to show low school performance. The sample statistics, including information about the gender of the study participants as well as their migration background, are displayed in Table 1. Analyses show that students from SEN-L schools were slightly older than students from Hauptschule,  $t(800) = 3.84, p < .001, d = 0.27$ . However, groups did not differ significantly with regard to gender,  $\chi^2 = 2.75, p = .10$ , to the proportion of students with migration background,  $\chi^2 = 0.20, p = .66$ , or to the proportion of students with German as their first language,  $\chi^2 = 3.46, p = .06$ . The proportion of female students in both types of schools was higher than in the respective German school types on average (36 % in schools for students with SEN-L or 46 % in Hauptschule; Statistisches Bundesamt 2012).

Students participated in the study voluntarily with personal and parental consent. Each participating student received a monetary incentive of 5 Euros for every testing session.

---

<sup>5</sup> During the early funding phase of NEPS, a series of feasibility studies was set up to experimentally test whether and, if so, how, students with SEN-L could be meaningfully included in the NEPS survey. Therefore, in contrast to data from regular school students, not all data from this group of students is available in the scientific use files of the NEPS.

<sup>6</sup> Students were identified as having special educational needs by the type of school they attended. An individual diagnostic process to identify students would have been too costly in time and money.

Table 1: Sample statistics of the study gathered in Grade 5

School type	<i>N</i>	Age [years] <i>M (SD)</i>	Female participants [%]	Students without migration background [%]	Students with German as first language [%]
SEN-L schools	407	11.40 (0.62)	46.9	71.9	82.3
Hauptschule	397	11.24 (0.59)	52.9	70.4	76.9

Information on migration/language background was available for 83 / 92 % of students with SEN-L and for 91 / 98 % of students attending Hauptschule only (see Nusser et al. 2013 for the validity of survey data of students with educational needs)

## Design

In order to investigate the effects of testing accommodations on the assessment of metacognitive knowledge, two conditions that varied in terms of administration mode were compared: *autonomous reading* similar to those in regular test settings and a *read-aloud* condition. Hence, a 2 (school type)  $\times$  2 (administration mode) design was implemented (see Table 2).

## Measures and procedures

Students participating in the study were surveyed with several competence and ability tests as well as questionnaires. Questionnaires contain information on gender, age, and migration background. These data as well as basic nonverbal cognitive skills and reading speed were assessed while students were in Grade 5. The metacognitive knowledge test was implemented 1 year later in Grade 6.

Prior to the current study, a series of qualitative pre-studies (individual interviews as well as group sessions) were conducted, in particular involving students with SEN-L. These pre-studies showed that the tests and questionnaires seem to imply a standard that – in principle – can be met by students with SEN-L. In addition, no avoidance behavior of students with SEN-L was observed in these qualitative pre-studies –

neither in individual interviews, nor in group testing situations.

All instruments used in the study were administered in groups of 10 to 20 students in classroom settings at their respective schools. Data were collected by the International Association for the Evaluation of Educational Achievement (IEA) Data Processing and Research Center (DPC) in Hamburg, Germany. All participating students completed the same tests and questionnaires. The testing procedure has been set up to be as comparable as possible to the procedure implemented in the regular NEPS samples. That is, the paper-and-pencil tests were administered in group settings and testing time was the same for all students. Students were introduced to the item format of the different tests via instructions and example items. In order to make sure that the students understood the given instructions and that the tests would be applied correctly, the test administrator interacted with the students to explain the item format on a poster and to find the correct solutions of the example items (e.g., “what do you think is the correct solution to this question?” ... “OK, then we have to tick this box right here”). This approach was applied to both groups of students and differed slightly from the procedure in regular NEPS test settings where students have to respond to the example items all by themselves.

Table 2: Design for the assessment of metacognitive knowledge

Administration mode	Students with SEN-L attending special schools	Students attending Hauptschule (lowest track of secondary education)
Autonomous reading	Regular testing situation	Regular testing situation
Read-aloud	Accommodated testing situation	Accommodated testing situation

## Reasoning

Reasoning as a basic nonverbal cognitive skill was assessed by the NEPS reasoning test (NEPS-MAT) developed by Lang et al. (2014). It was designed as a matrices test in the tradition of the Raven test (Raven et al. 2003). Each item consists of several horizontally and vertically arranged fields, in which different geometrical elements are shown – with only one field remaining blank. Students were asked to select the right

complement for the blank field from the offered solutions. In order to do this, the logical rules according to which the pattern of geometrical elements has been arranged must be deduced. The test consists of three subsets with four items each. Students were given nine minutes (3  $\times$  3 min) to solve the items. Performance was rated by the sum of correctly solved items (for more information, see Haberkorn and Pohl 2013).

### Reading speed

In order to assess reading speed, a test was implemented on the basis of the principles of the *Salzburger Lesescreening* (Auer et al. 2005). The test consists of 51 sentences presented in order of sentence length, with sentences ranging from 5 to 18 words. Students were asked to read as many sentences as possible and to judge their truth (true vs. false; sample item: “There is a bath tub in every garage”). Testing time was limited to 2 min. Students were scored in terms of the sum of correctly judged sentences. For more detailed information on the test, see Zimmermann et al. (2012).

### Metacognitive knowledge

Metacognitive knowledge was measured by a scenario-

based procedure focusing primarily on different aspects of strategy knowledge. The test consisted of eight scenarios describing different school and leisure-time activities. More specifically, five of the scenarios were related to a school or learning context (two in the domain of reading), whereas the remaining three scenarios were embedded in out-of-school contexts, asking for domain-general strategy knowledge (see Händel et al. 2013). For each scenario six strategies (i.e., 48 strategies in total) were provided that varied in their degree of effectiveness for the given situation (cf., Händel et al. 2013). The strategies in the scenarios refer to a broad range of strategies, including cognitive, metacognitive, and resource management strategies. Subjects were asked to rate the effectiveness of each strategy on a 4-point scale of usefulness. A sample scenario is given in Fig. 1.

Peter has a lot to do this week: He is supposed to go to the swimming club twice, he has been given plenty of homework, and he has to buy a birthday present for his friend.

What should he do in order to manage everything?

*Please judge the usefulness of the proposed strategies.*

	not useful at all	barely useful	somewhat useful	very useful
He makes a plan for the week and organizes his time for the tasks. He follows his plan very closely.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
He combines different tasks and buys the birthday present on his way to the swimming club.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
He allows others to help him. He asks his brother to buy the birthday present.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
He completes only those pieces of homework which can be done quickly. Then he deals with the other things.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
First, he buys the birthday present. If this takes too long, he will skip the homework or the swimming.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
He does exactly what he feels like doing at this moment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 1: Sample scenario of the test about metacognitive knowledge. Note that the displayed scenario has not been included in the final NEPS test

As outlined earlier, two different administration modes were implemented in both samples: the standard test situation and a read-aloud condition. In the standard condition (autonomous reading), students had to read through the metacognitive knowledge test and judge the alternate strategies on their own. In the read-aloud condition students were guided through the test verbally. In the latter condition, a CD was played to the students, containing all the instructions, scenarios, and

strategy alternatives read aloud. Pauses of 1 min after each of the eight different scenarios were included, allowing students to rate the strategy alternatives. In both administration modes, students were asked not to tick the boxes of the rating scale for each proposed approach before all approaches to a given scenario had been read. Testing time was 15 min in both administration modes.

To develop an objective scoring procedure for the

students' responses, scientists in the field of educational psychology and learning strategies were asked to provide their judgments on the appropriateness of each strategy. These experts' judgments of the relative usefulness of the presented alternatives (e.g., higher rating for strategy A than for strategy B) were used as a reference for the scoring of the test. That is, a pair comparison was scored as correct if the judgment on a strategy pair was in line with the experts' ratings, and as incorrect if the judgment on a strategy pair was contrary to the experts' ratings, or if the two strategies of a pair were considered as equal (for further information on this procedure, see Händel et al. 2013; Lockl 2012). Overall, the test includes 69 pair comparisons, with Cronbach's  $\alpha = .85/.89$  for students with SEN-L/students attending Hauptschule. The reported score represents an overall mean test score that refers to the proportion of correct pair comparisons in relation to all pair comparisons the students had completed. The values of the mean test score ranged from 0 (no pair comparisons solved correctly) to 100 % (all pair comparisons solved correctly).

## Results

We first report descriptive statistics on the control variables and on the metacognitive knowledge test. Thereafter, we refer to results of the analyses of variance (ANOVAs) to examine the effects of school type and testing accommodations. Finally, correlations between the scores on the metacognitive knowledge test with reading speed and reasoning ability are reported.

### Control variables

The descriptive statistics for the assessed control variables are given in Table 3. As expected, reasoning and reading speed appeared to be rather low for both groups of students in comparison to the average scores in the representative NEPS main sample of students in fifth grade ( $M = 6.97$ ,  $SD = 2.60$  for reasoning and  $M = 21.24$ ,  $SD = 7.00$  for reading speed). Nevertheless, students from Hauptschule clearly surpassed students from SEN-L schools.

Table 3: Statistics of reasoning as well as reading speed for students with SEN-L and students attending Hauptschule

Variable	Scale max.	SEN-L school	Hauptschule	$t$ (796)	$d$
Reasoning	12	3.43 (2.17)	5.93 (2.48)	14.69***	1.07
Reading speed	51	11.14 (5.61)	18.87 (6.53)	17.97***	1.27

\*\*\*:  $p < .001$

### Metacognitive knowledge test

With regard to the overall level of the ratings that students provided for the different strategies, the following results emerged: Students with SEN-L generally showed higher ratings on the 4-point rating scale ( $M = 2.83$ ,  $SD = 0.38$ ) than students attending Hauptschule ( $M = 2.59$ ,  $SD = 0.25$ ,  $t(802) = 10.53$ ,  $p < .001$ ,  $d = 0.75$ ) – that is, students with SEN-L generally estimated strategies to be more useful than students attending Hauptschule. Furthermore, we examined the mean ratings for useful and not useful strategies. Out of the 48 strategies, 20 were specified as useful (i.e., no other strategy within a given scenario was rated as more useful by the experts) and 16 as not useful strategies (i.e., no other strategy within this scenario was rated as less useful by the experts). Concerning the useful strategies, the mean ratings were significantly lower for students with SEN-L ( $M = 2.95$ ,  $SD = 0.44$ ) than for students attending Hauptschule ( $M = 3.10$ ,  $SD = 0.35$ ,  $t(802) = 5.29$ ,  $p < .001$ ,  $d = 0.38$ ). Conversely, for strategies defined as not useful, students with SEN-L provided higher ratings ( $M = 2.65$ ,  $SD = 0.46$ ) than students attending Hauptschule ( $M = 2.02$ ,  $SD = 0.39$ ,  $t(802) = 21.83$ ,  $p < .001$ ,  $d = 1.48$ ). The effect sizes reveal that the difference between the students from the two school types is considerably more pronounced for not useful strategies than for useful strategies. Hence, students with SEN-L seem to be less able to detect strategies that

are not useful. In addition, students with SEN-L do not differentiate very much between useful and not useful strategies, although the difference is still significant,  $t(406) = 12.67$ ,  $p < .001$ ,  $d = 0.67$ .

Regarding missing values, our data indicate that students, in general, showed only few missing values in the ratings on the metacognitive knowledge test ( $M = 1.20$ ,  $SD = 4.40$ ), indicating that they did not refuse to take part in the test. However, students with SEN-L showed more missing values ( $M = 2.05$ ,  $SD = 5.99$ ) than students attending Hauptschule ( $M = 0.33$ ,  $SD = 1.04$ ,  $t(802) = 5.64$ ,  $p < .001$ ,  $d = 0.40$ ).

Next, we investigated the test scores on the basis of pair comparisons in the metacognitive knowledge test. On average, students received a mean score of 55.63 % ( $SD = 17.87$ ) in the metacognitive knowledge test – that is, they were able to solve about half of the pair comparisons. Before analyzing the effects of the testing accommodations, we examined whether the score on metacognitive knowledge differed as a function of gender, first language, or migration background. Whereas no significant differences were found for gender,  $t(798) = 1.79$ ,  $p > .05$ ,  $d = 0.12$ , and first language,  $t(759) = 1.42$ ,  $p > .05$ ,  $d = 0.13$ , differences were detected due to migration history: students without migration background ( $M = 56.8\%$ ,  $SD = 18.2$ ) outperformed students with migration background ( $M = 53.4\%$ ,  $SD = 17.7$ ) on the metacognitive knowledge test,  $t(696) = 2.24$ ,  $p < .05$ ,  $d = 0.19$ .

However, the effect size is small. As a consequence, and because these demographic variables were not available for the whole sample, we did not take them into consideration in our further analyses.

### Effects of school type and testing accommodations on metacognitive knowledge

To test for group differences regarding the metacognitive knowledge score, analyses of variance were conducted with the metacognitive knowledge test score as dependent variable and school type and administration mode as independent variables. The corresponding descriptive statistics are displayed

in Table 4.

The effect of school type was statistically significant and had a high effect size,  $F(1, 797) = 508.99, p < .001, \eta_p^2 = .39$ . By contrast, the effect of the administration mode (autonomous reading vs. read-aloud) – although significant – was negligible,  $F(1, 797) = 8.77, p < .005, \eta_p^2 = .01$ . The interaction effect of school type and reading condition ( $F(1, 797) = 6.14, p < .05, \eta_p^2 = .01$ ) indicated that the reading condition effect was evident for students from Hauptschule (simple effect:  $F(1, 797) = 19.73, p < .001$ ) but not for students from SEN-L schools (simple effect:  $F(1, 797) = 3.11, p = .08$ ; see Fig. 2).

Table 4: Descriptive statistics of the metacognitive knowledge test score

School type	Condition	Metacognitive knowledge (%)
SEN-L	Autonomous reading	43.66 (12.92)
	Read-aloud	44.18 (13.66)
	Total	43.89 (13.24)
Hauptschule	Autonomous reading	63.66 (16.09)
	Read-aloud	68.90 (13.29)
	Total	66.38 (14.92)
Total Sample	Autonomous reading	53.78 (17.37)
	Read-aloud	57.52 (18.20)

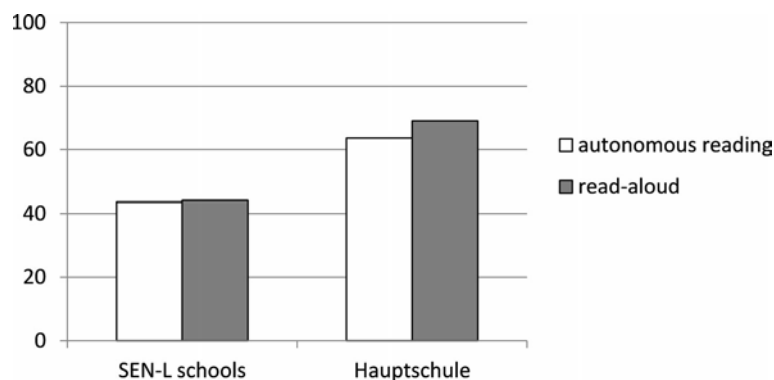


Fig. 2: Metacognitive knowledge score for school type (SEN-L and Hauptschule) by administration mode (autonomous reading vs. read-aloud)

We also examined whether the testing accommodation had any effect on the number of missing values. An analysis of variance revealed a significant effect of school type mentioned previously ( $F(1, 800) = 29.52, p < .001, \eta_p^2 = .04$ ), a significant main effect of administration mode ( $F(1, 800) = 5.52, p < .05, \eta_p^2 = .01$ ), and a significant interaction school type by administration mode ( $F(1, 800) = 4.92, p < .05, \eta_p^2 = .01$ ). That is, the number of missing values was higher in the autonomous reading condition ( $M = 1.60, SD = 5.47$ ) than in the read-aloud condition ( $M = 0.77, SD = 2.82$ ). The interaction effect indicated that students with SEN-L benefited from the read-aloud condition in terms of fewer missing values ( $M = 1.29,$

$SD = 3.86$  in the read-aloud condition and  $M = 2.86, SD = 7.24$  in the autonomous reading condition). However, this was not true for students from Hauptschule ( $M = 0.35, SD = 0.91$  in the autonomous reading condition and  $M = 0.31, SD = 1.15$  in the read-aloud condition, respectively).

In sum, the results indicate that a read-aloud test accommodation leads to fewer missing values and higher test scores (first research question of the study). Regular school students provided higher test scores than students with SEN-L. In addition, the read-aloud accommodation assisted especially the performance of regular students. In students with SEN-L the reduced



number of missing values did not result in better performance scores (research question 2).

### Interrelations with reasoning and reading speed

To detect the possible influence of factors on the valid assessment of metacognitive knowledge, correlations with reasoning and reading speed were calculated. Moderate correlations were found. As can be seen in Table 5, the strength of the correlations differed as a function of the administration mode. The correlations in both school types were slightly higher in the autonomous reading condition than in the read-aloud condition. However, the difference between the

administration modes became more apparent with regard to reading speed. In the read-aloud condition non-significant correlations between metacognitive knowledge and reading speed emerged in contrast to significant correlation coefficients in the autonomous reading condition. That is, when students had to read the scenarios and corresponding strategies on their own, the scores on metacognitive knowledge were significantly associated with their reading speed, whereas this was not the case when the test was read aloud to the students. Furthermore, the difference between the correlation coefficients in the two conditions was significant for students attending Hauptschule ( $z = 4.19, p < .01$ ) but not for the students with SEN-L ( $z = 0.67, p > .05$ ).

Table 5 Correlation coefficients of metacognitive knowledge and reasoning and reading speed for the different groups of students

School type	Condition	Reasoning	Reading speed
SEN-L	Autonomous reading	.20**	.15*
	Read-aloud	.17*	.08
	Total	.19**	.12*
Hauptschule	Autonomous reading	.30**	.39**
	Read-aloud	.25**	-.01
	Total	.27**	.19**
Total Sample	Autonomous reading	.47**	.50**
	Read-aloud	.44**	.37**

\*\* $: p < .01$ , \* $: p < .05$

As a consequence, we split the sample by the median value of reading speed. In order to produce similar cell sizes and as the reading speed score significantly differed between the two school types, we computed the median split within each school type ( $Md = 11$  for students with SEN-L and  $Md = 18$  for students attending Hauptschule).<sup>7</sup> An ANOVA with the median split, school type, and condition as independent variables and metacognitive knowledge as dependent variable resulted in the following effects: Besides significant main effects due to school type and administration mode (already reported in Fig. 2), a significant effect due to reading speed occurred: Fast readers surpassed slow readers,  $F(1, 787) = 26.96, p < .001, \eta_p^2 = .04$ . In addition, the two-way interaction effects of school type by administration mode ( $F(2, 787) = 5.13, p < .05, \eta_p^2 = .01$ ) and of reading speed by administration mode ( $F(1, 787) = 3.78, p < .05, \eta^2 = .01$ ) proved to be significant. Finally, there was a significant three-way interaction ( $F(1, 787) = 4.05, p < .05, \eta^2 = .01$ ), showing that the effect of the administration mode (autonomous reading vs. read-aloud) was only evident for slow readers attending Hauptschule (cf., Fig. 3).

The reported results indicate that reading speed significantly influenced test performance in a metacognitive knowledge test (research question 3).

## Discussion

The valid assessment of metacognitive knowledge is of great relevance for research and education. To educate students and to teach them to become successful learners, it is fundamental to have information about the strategy knowledge that students are equipped with, to know whether they use strategies spontaneously, and if not, to know why they do not. With respect to students with SEN-L, the assessment of metacognitive competencies is of special importance. First, students with SEN-L are assumed to lack metacognitive competencies that are required to regulate one's own learning. Second, in Germany, the majority of all students who leave school without a regular qualification certificate after the period of compulsory school attendance have attended a special school for students with SEN-L (Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany 2012).

<sup>7</sup> The definition of fast and slow in this context is a relative one within each reference group. Fast readers who were members of the group of students with SEN-L might still be slower readers than slow readers attending Hauptschule.

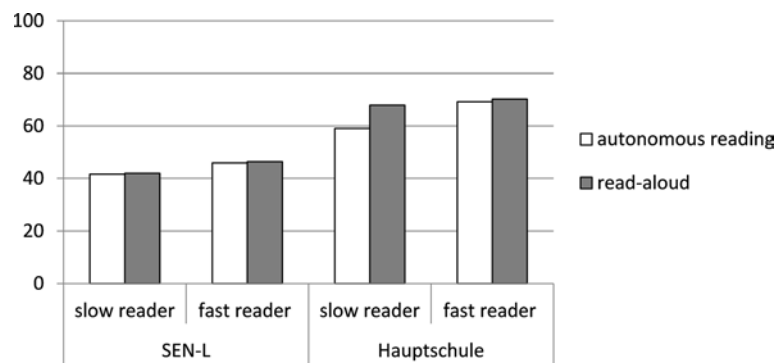


Fig. 3 Metacognitive knowledge score for students at SEN-L schools and Hauptschule, differing in reading speed (slow vs. fast), and interacting with administration mode (autonomous reading vs. read-aloud)

From a research perspective, assessment instruments should be valid and reliable. In addition, large-scale studies need to reduce costs of administration (group settings where possible) and analysis (closed answer formats that allow for numerical coding) (Blossfeld et al. 2011). Considering these features is a challenging endeavor for the assessment of students with SEN-L. Feasibility studies within the NEPS met this challenge by oversampling students with SEN-L from special schools. The reported feasibility study of the NEPS surveyed students with SEN-L as well as a group of students who were most closely comparable to students with SEN-L – that is, students from the same grade attending regular schools from the lowest track of secondary education. Data concerning the sample characteristics show that the sampling procedure was successful, that is, both groups of students were quite comparable regarding demographic variables, such as gender and age as well as first language and migration background. However, and as expected, students of Hauptschule had considerably higher test scores in the reasoning as well as in the reading speed test.

Metacognitive knowledge was assessed via a paper-and-pencil instrument that was suitable for the group-administered test settings implemented in the NEPS. The scores of the metacognitive knowledge test indicated an appropriate difficulty level and the test could be shown as being reliable in both samples. In accordance with our expectations, we found that students with SEN-L displayed lower levels of metacognitive knowledge in comparison to students attending Hauptschule. In fact, the difference in performance between the two groups of students is considerable. This result can be interpreted in two different ways, which are not mutually exclusive. First, the low performance of students with SEN-L in the metacognitive knowledge test could be seen as an actual deficit in metacognitive knowledge. This interpretation is consistent with previous research (e.g., Paris and Oka 1986; Pintrich et al. 1994) showing that students with learning disabilities lack metacognitive knowledge. Second, there is the possibility that other construct-irrelevant disabilities may have influenced the test score in the metacognitive knowledge test in students with SEN-L (Pitoniak and Royer 2001; Sireci et al. 2005).

This second explanation was investigated further with regard to reading abilities as an assumed construct-irrelevant ability that might have hampered test processing. Therefore, two different administration modes (read-aloud compared to autonomous reading) were implemented in both groups of students. The results concerning this testing accommodation revealed that the read-aloud administration of the metacognitive knowledge test facilitated test processing in comparison to the autonomous reading condition. However, more detailed analyses showed that this was only evident for students attending Hauptschule. This is an unexpected finding as the read-aloud accommodation of the administration was originally designed for students with SEN-L, which thus contradicts the differential boost hypothesis (Cormier et al. 2010). In fact, the read-aloud administration led to fewer missing values in students with SEN-L but not to higher test scores itself. We assume that the oral accommodation supported students to engage in the test. However, reading-aloud did not improve the test scores of students with SEN-L.

Additional correlation analyses indicated a significant positive relationship between students' reading speed and their metacognitive knowledge score. Interestingly, the correlation scores were higher for students in the autonomous reading condition than in the read-aloud condition. As long as students had to read on their own without oral support, reading speed – a construct-irrelevant ability – might have influenced the metacognitive knowledge test score. Accordingly, the pattern of results suggests that the read-aloud administration allowed for a more valid assessment of metacognitive knowledge.

Subsequent analyses with reading speed as a potential moderating variable highlighted that only slow readers attending Hauptschule benefited from the read-aloud administration, that is, they received higher metacognitive knowledge scores compared to students in the autonomous reading condition. In all other groups, differences between administration modes were negligible. It seems that more proficient readers attending Hauptschule did not need the testing accommodation and that they were able to work on the test themselves without any additional facilities (Sireci et al. 2005). In contrast, and quite surprisingly, the oral

accommodation was not supportive of students with SEN-L, neither of the less nor of the more proficient readers. As a consequence, the questions arise why do students with SEN-L display rather low levels of metacognitive knowledge and why do they not benefit from the testing accommodation? Looking at the pattern of results, it seems as if deficits in reading skills were not the main reason for the low scores in the metacognitive knowledge test. Moreover, both slow and fast readers in the group of students with SEN-L still were correct on more than 40 % of the pair comparisons and showed good test reliability. It seems unlikely that students could not even follow the oral information within their booklets, which would have resulted in a relatively random selection of the strategy alternatives and lower test scores.

Instead, the results suggest that students with SEN-L experienced severe difficulties judging the usefulness of the presented strategies and also that they struggled to clearly distinguish between more or less successful strategies. For example, students with SEN-L rated strategies such as “he/she is confident in remembering the steps”, or “he/she relies on the belief that she took care of everything” as relatively useful for remembering several things or steps. Accordingly, students with SEN-L, on average, provided high ratings, even for strategies that were not considered to be useful in this scenario but might sound useful in a more general situation. Thus, it appears that students with SEN-L do not think through the scenarios thoroughly but judge the provided strategies in a rather superficial way.

This seems to indicate that students with SEN-L actually do lack metacognitive knowledge. Nevertheless, it cannot be ruled out that further construct-irrelevant disabilities may be at least partially responsible for the low performance of students with SEN-L. Possibly, the memory or attention span of students with SEN-L (Abedi et al. 2010) could have been too restricted to remember the situation according to which they should judge the strategy options after having been provided with six different strategic approaches. Accordingly, one possibility is that students did not rate the approaches provided with regard to the usefulness of the linked scenario but rather in relation to their overall perception of whether this seemed like a reasonable strategy in itself.

## Limitations of the study

One limitation of the present study is that no other measures of metacognition, such as metacognitive regulation processes, were assessed. That is, the current study does not provide any information about the actual strategic behavior and its assessment in students with SEN-L. Metacognitive knowledge about strategies can be considered as a precondition for the successful regulation of learning behavior. However, high metacognitive knowledge about the usefulness of certain strategies in specific situations does not lead to the corresponding strategic behavior per se – that is, the application of this knowledge in the actual learning situation. Previous research has shown that students

with learning disabilities exhibit problems regarding the regulation of learning (Meltzer et al. 2004; Schröder 2000), and so measuring procedural aspects of metacognition appears worthwhile. However, it has to be acknowledged that within large-scale studies, it is difficult to assess regulation aspects. Hence, measures for error detection or allocation of study time (Lockl and Schneider 2003; Metcalfe 2009; Pieschl 2009), or thinking-aloud protocols (Veenman 2005) typically require measurements of students’ ongoing learning processes and need individual testing of the students which was not possible within the NEPS.

The experimental study is limited because only one specific accommodation (administration mode of reading-aloud) has been implemented and therefore the effects of other test accommodations need to be further investigated. If, in fact, students’ working memory is a limiting factor, then a reduction in the number of presented alternatives per situation might lead to higher test scores. Test instruments with only three strategies per scenario are implemented in the NEPS studies in Grade 1 and Grade 3, for example. In the first grade the scenarios and proposed strategies are additionally accompanied by pictures that also function as a memory aid (Lockl et al. 2013).

Finally, as the sample of students with SEN-L attending special schools is expected to be heterogeneous, the accommodation might have only had a positive effect for some but not for all students (compare Sireci et al. (2005) who showed that oral accommodations on math tests were associated with increased test performance for some but not all students with disabilities). The high standard deviation of the number of missing values within the group of students attending SEN-L schools indicates that some students were having difficulties in processing the test. Finally, due to the recruitment procedure used in the study, it cannot be ruled out that some of the students recruited from Hauptschule may also have had special educational needs but that they may not have been identified as such yet. Students with difficulties or students who are underachieving usually pass through a diagnostic process before they are advised to attend special schools. Nevertheless, there is some overlap between students attending the lowest track of secondary education in Germany and students attending special schools. In addition, students with SEN-L and students with language impairments, emotional, or social problems may be educated at the same type of school (Bos et al. 2010). However, the results were quite unequivocal with regard to school type so that this should not have influenced the results in a substantial way.

## Assessment implications and future directions

The study carries implications for the assessment of students with SEN-L. The implemented test accommodation did not lead to higher test scores in students with SEN-L per se. Instead, when students with SEN-L complete a specific test, the suitability of the testing characteristics have to be carefully checked.

Hence, further research is called for to examine whether some students with SEN-L might also lack other abilities besides language/reading that would enable them to successfully complete the metacognitive knowledge test. At least, the students did not refuse to engage in the test and showed only few missing values and, therefore, seemed to be able to complete the test in general. The metacognitive knowledge test showed high internal consistency in both groups of students and even students with SEN-L were able to significantly differentiate between useful and not useful strategies – although the magnitude of this difference was comparably small. That is, the newly developed scenario-based metacognitive knowledge test allowed for the assessment of metacognitive knowledge within a group setting. Nevertheless, additional analyses are needed to identify which students were having notable difficulties and if other accommodations would have led to higher test scores. Inter alia, qualitative data analyses are needed to investigate the response behavior of students with SEN-L in more detail – that is, whether or not these students rated specific types of strategies as very useful even though they were not.

## References

- Abedi, J., Kao, J. C., Leon, S., Mastergeorge, A. M., Sullivan, L., Herman, J., et al. (2010). Accessibility of segmented reading comprehension passages for students with disabilities. *Applied Measurement in Education*, 23, 168–186. doi:10.1080/08957341003673823.
- Artelt, C., Beinicke, A., Schlagmüller, M., & Schneider, W. (2009). Diagnose von strategiewissen beim textverstehen [diagnosing strategic knowledge in text comprehension]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 41(2), 96–103. doi:10.1026/0049-8637.41.2.96.
- Auer, M., Gruber, G., Mayringer, H., & Wimmer, H. (2005). *Salzburger Lesescreening für die Klassenstufen 5–8 [Salzburg reading screening for Grades 5–8]*. Göttingen: Hogrefe.
- Barkow, I., Leopold, T., Raab, M., Schiller, D., Wenzig, K., & Blossfeld, H.-P. (2011). RemoteNEPS: Data dissemination in a collaborative workspace. In H.-P. Blossfeld, H.-G. Roßbach, & J. v. Maurice (Eds.), *Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 315–325). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Belmont, J. M., & Borkowski, J. G. (1988). A group-administered test of children's metamemory. *Bulletin of the Psychonomic Society*, 26(3), 206–208.
- Blossfeld, H.-P., & von Maurice, J. (2011). Education as a lifelong process. In H.-P. Blossfeld, H.-G. Roßbach, & J. V. Maurice (Eds.), *Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 19–34). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Blossfeld, H.-P., Maurice, J. V., & Schneider, T. (2011). The National Educational Panel Study: need, main features, and research potential. *Zeitschrift für Erziehungswissenschaft*, 14, 5–17. doi:10.1007/s11618-011-0178-3.
- Bolt, S. E., & Ysseldyke, J. (2007). Accommodating students with disabilities in large-scale testing: a comparison of differential item functioning (DIF) identified across disability types. *Journal of Psychoeducational Assessment*, 26, 121–138. doi:10.1177/0734282907307703.
- Bos, W., Müller, S., & Stubbe, T. C. (2010). Abgehängte Bildungsinstitutionen: Hauptschulen und Förderschulen [Educational institutions left behind: The lowest track of secondary education (Hauptschulen) and special schools]. In K. Hurrelmann & G. Quenzel (Eds.), *Bildungsverlierer: Neue Ungleichheiten* (pp. 375–398). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bosson, M. S., Hessels, M. G. P., Hessels-Schlatter, C., Berger, J.-L., Kipfer, N. M., & Büchel, F. P. (2010). Strategy acquisition by children with general learning difficulties through metacognitive training. *Australian Journal of Learning Difficulties*, 15, 13–34. doi:10.1080/19404150903524523.
- Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.) *Advances in instructional psychology* (Vol. 1, pp. 77–165). Mahwah, NJ: Erlbaum.
- Bundesamt, S. (2012). *Schulen auf einen Blick [schools at a glance]*. Wiesbaden: Statistisches Bundesamt.
- Cormier, D. C., Altman, J., Shyyan, V., & Thurlow, M. L. (2010). A summary of the research on the effects of test accommodations: 2007–2008. (Vol. 56). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Elbaum, B., Arguelles, M. E., Campbell, Y., & Saleh, M. B. (2004). Effects of a student-reads-aloud accommodation on the performance of students with and without learning disabilities on a test of reading comprehension. *Exceptionality*, 12(2), 71–87. doi:10.1207/s15327035ex1202\_2.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist*, 34, 906–911.
- Fletcher, J. M. (2012). Classification and identification of learning disabilities. In B. Y. L. Wong & D. L. Butler (Eds.), *Learning about learning disabilities* (4th ed., pp. 1–25). Amsterdam: Elsevier Academic Press.
- Fuchs, L. S., & Fuchs, D. (1999). Fair and unfair testing accommodations. *The School Administrator*, 56, 24–27. Haberkorn, K., & Pohl, S. (2013). *Cognitive basic skills – Data in the scientific use file*. Bamberg: Otto-Friedrich-University, National Educational Panel Study.
- Hammill, D. D. (1990). On defining learning disabilities: an emerging consensus. *Journal of Learning Disabilities*, 23, 74–84. doi:10.1177/002221949002300201.
- Händel, M., Artelt, C., & Weinert, S. (2013). Assessing

- metacognitive knowledge: development and evaluation of a test instrument. *Journal of Educational Research Online*, 5, 162–188.
- Hannah, C. L., & Shore, B. M. (1995). Metacognition and high intellectual ability: insights from the study of learning-disabled gifted students. *Gifted Child Quarterly*, 39, 95–109. doi:10.1177/001698629503900206.
- Hasselhorn, M. (1994). Zur Erfassung von Metagedächtnisaspekten bei Grundschulkindern [On the assessment of metacognitive aspects in elementary school children]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 26, 71–78.
- Heydrich, J., Weinert, S., Nusser, L., Artelt, C., & Carstensen, C. H. (2013). Including students with special educational needs into large-scale assessments of competencies: challenges and approaches within the German National Educational Panel Study (NEPS). *Journal for Educational Research Online*, 2, 217–240.
- Hollenbeck, K. (2002). Determining when test alterations are valid accommodations or modifications for large-scale assessment. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation* (pp. 395–425). Mahaw: Lawrence Erlbaum.
- Hörmann, B. (2007). *Die Unsichtbaren in PISA, TIMMS, und Co [The invisibles in PISA, TIMMS, and Co]*. Vienna: University of Vienna.
- Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about reading: issues in definition, measurement, and instruction. *Educational Psychologist*, 22(3/4), 255–278.
- Johnson, E. S. (2000). The effects of accommodations on performance assessments. *Remedial and Special Education*, 21(5), 261–267. doi:10.1177/074193250002100502.
- Ketterlin-Geller, L. R., Alonzo, J., Braun-Monegan, J., & Tindal, G. (2007). Recommendations for accommodations: implications of (in)consistency. *Remedial and Special Education*, 28, 194–206. doi:10.1177/07419325070280040101.
- Koretz, D., & Barton, K. (2003). Assessing students with disabilities: issues and evidence. *CSE Technical Report 587*. Los Angeles: CSE Technical Report 587.
- Körkel, J. (1987). *Die Entwicklung von Gedächtnis- und Metagedächtnisleistungen in Abhängigkeit von bereichsspezifischen Vorkenntnissen [The development of memory and metamemory performance as dependent on domain-specific prior knowledge]*. Frankfurt: Peter Lang.
- Kreutzer, M. A., Leonard, C., & Flavell, J. H. (1975). An interview study of children's knowledge about memory. *Monographs of the Society for Research in Child Development*, 40.
- Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens [psychological diagnostic: theory and praxis of psychologic diagnosis]*. Wien: Hogrefe.
- Kurtz, B. E., Reid, M. K., Borkowski, J. G., & Cavanaugh, J. C. (1982). On the reliability and validity of children's metamemory. *Bulletin of the Psychonomic Society*, 19, 137–140.
- Laitusis, C. C. (2010). Examining the impact of audio presentation on tests of reading comprehension. *Applied Measurement in Education*, 23, 153–167.
- Lang, F. R., Kamin, S., Rohr, M., Stünkel, C., & Williger, B. (2014). *Erfassung der fluiden kognitiven Leistungsfähigkeit über die Lebensspanne im Rahmen des Nationalen Bildungspanels: Abschlussbericht zu einer NEPS-Ergänzungsstudie [Assessment of fluid intelligence across the life span within the German National Educational Panel Study: Final report about the supplementary study]* (NEPS Working Paper No. 43). Bamberg: Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel.
- Lehmann, R., & Hoffmann, E. (2009). *BELLA. Berliner Erhebung arbeitsrelevanter Basiskompetenzen von Schülerinnen und Schülern mit Förderbedarf „Lernen“*. [BELLA. Berlin survey of work relevant basic competencies of students with SEN-L]. Münster: Waxmann.
- Lockl, K. (2012). *Assessment of declarative metacognition: Starting cohort 4 – ninth grade*. Bamberg: Otto-Friedrich-University, Nationales Bildungspanel.
- Lockl, K., & Schneider, W. (2003). Metakognitive Überwachungs- und Selbstkontrollprozesse bei der Lernzeiteinteilung von Kindern [Metacognitive processes of monitoring and self-control in allocating children's study time]. *Zeitschrift für Pädagogische Psychologie*, 17, 173–183. doi:10.1024//1010-0652.17.3.173.
- Lockl, K., Händel, M., Haberkorn, K., & Weinert, S. (2013). Metacognitive knowledge in young children: Development of a new test procedure for first graders. In H.-P. Blossfeld, J. v. Maurice, & J. Skopek (Eds.), *Methodological issues of longitudinal surveys: The example of the National Educational Panel Study*. Manuscript submitted for publication.
- Maag Merki, K., Ramseier, E., & Karlen, Y. (2013). Reliability and validity analyses of a newly developed test to assess learning strategy knowledge. *Journal of Cognitive Education and Psychology*, 12, 391–408.
- McDonnell, L. M., McLaughlin, M., & Morrison, P. (1997). *Educating one and all: Students with disabilities and standards-based reform*. Washington: National Academy Press.
- Meltzer, L., Katzir, T., Miller, L., Reddy, R., & Roditi, B. (2004). Academic self-perceptions, effort, and strategy use in students with learning disabilities:

- changes over time. *Learning Disabilities Research & Practice*, 19, 99–108. Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, 18, 159–163. doi:10.1111/j.1467-8721.2009.01628.x.
- Myers, M., & Paris, S. G. (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology*, 70, 680–690.
- NCLD – National Center for Learning Disabilities. (2007). *State testing accommodations: A look at their value and validity*. Retrieved from <http://www.cehd.umn.edu/nceo/OnlinePubs/NCLD/NCLDStateTestingAccommodationsStudy.pdf>
- Neuenhaus, N., Artelt, C., Lingel, K., & Schneider, W. (2011). Fifth graders metacognitive knowledge: general or domain-specific? *European Journal of Psychology of Education*, 26, 163–178. doi:10.1007/s10212-010-0040-7.
- Nusser, L., Heydrich, J., Carstensen, C. H., Artelt, C., & Weinert, S. (2013). Validity of survey data of students with special educational needs – Results from the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, & J. Skopek (Eds.), *Methodological issues of longitudinal surveys. The example of the National Educational Panel Study*. Manuscript submitted for publication.
- Paris, S. G., & Oka, E. R. (1986). Children's reading strategies, metacognition, and motivation. *Developmental Review*, 6, 25–56.
- Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8, 293–316.
- Pieschl, S. (2009). Metacognitive calibration – an extended conceptualization and potential applications. *Metacognition and Learning*, 4, 3–31. doi:10.1007/s11409-008-9030-4.
- Pintrich, P. R., Anderman, E. M., & Klobucar, C. (1994). Intraindividual differences in motivation and cognition in students with and without learning disabilities. *Journal of Learning Disabilities*, 27, 360–370.
- Pitoniak, M. J., & Royer, J. M. (2001). Testing accommodations for examinees with disabilities: a review of psychometric, legal, and social policy issues. *Review of Educational Research*, 71, 53–104.
- Pressley, M., Borkowski, J. G., & Schneider, W. (1989). Good information processing: what it is and how education can promote it. *International Journal of Educational Research*, 13, 857–867.
- Raven, J., Raven, J. C., & Court, J. H. (2003). *Raven's progressive matrices und vocabulary scales*. Frankfurt: Pearson Assessment.
- Reid, R. R., Harris, K. R., Graham, S., & Rock, M. (2004). Self-regulation among students with LD and ADHD. In B. Wong & D. L. Butler (Eds.), *Learning about learning disabilities*. Amsterdam: Elsevier Academic Press. Sáez, L., Jamgochain, E., & Tindal, G. (2013). Accommodating special needs for large-scale assessments. In M. Simon, K. Ercikan, & M. Rousseau (Eds.), *Improving large-scale assessment in education* (pp. 125–140). New York: Routledge.
- Schlagmüller, M., & Schneider, W. (2007). *WLST 7–12. Würzburger Lesestrategie-Wissenstest für die Klassen 7 bis 12 [Würzburg reading strategy knowledge test for grades 7 through 12]*. Göttingen: Hogrefe.
- Schmitt, M. C. (1990). A questionnaire to measure children's awareness of strategic reading processes. *The Reading Teacher*, 43, 454–461.
- Schneider, W., & Pressley, M. (1997). *Memory development between 2 and 20*. Hillsdale: Erlbaum.
- Schröder, U. (2000). *Lernbehindertenpädagogik: Grundlagen und Perspektiven sonderpädagogischer Lernhilfe [Educating people with learning disabilities: Basics and perspectives of learning aids in special education]*. Stuttgart: Kohlhammer.
- Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: an analysis of the interaction hypothesis. *Review of Educational Research*, 75, 457–490. doi:10.3102/00346543075004457.
- Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (1999). *Empfehlungen zum Förderschwerpunkt Lernen [Recommendations according to special education support in learning]*.
- Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (2008). *Sonderpädagogische Förderung in Schulen 1997–2006 [Special education support in schools 1997–2006]*.
- Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (2012). *Sonderpädagogische Förderung in Schulen 2001 bis 2010 [Special education support in schools 2001 to 2010]*.
- Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology*, 82, 306–314.
- Swanson, H. L., & Trahan, M. (1996). Learning disabled and average readers' working memory and comprehension: does metacognition play a role? *British Educational Research Journal*, 66, 333–355.
- Thurlow, M. L. (2002). Accommodations for students with disabilities in high school. *National Center on Secondary Education and Transition*, 1(1), 1–6.
- Thurlow, M. L. (2010). Steps toward creating fully accessible reading assessments. *Applied Measurement in Education*, 23, 121–131. doi:10.1080/08957341003673765.
- Thurlow, M. L., Bremer, C., & Albus, D. (2008). *Good news and bad news in disaggregated subgroup reporting to the public on 2005–2006 assessment results (Technical Report 52)*. Minneapolis: University of Minnesota, National Center on

Educational Outcomes.

- Twomey, E. (2006). Linking learning theories and learning difficulties. *Australian Journal of Learning Disabilities*, 11, 93–98. doi:10.1080/19404150609546812.
- van Kraayenoord, C., & Schneider, W. E. (1999). Reading achievement, metacognition, reading self-concept and interest: a study of German students in grades 3 and 4. *European Journal of Psychology of Education*, 14, 305–324.
- Veenman, M. V. J. (2005). The assessment of metcognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und Metakognition. Implikationen für Forschung und Praxis [Learning strategies and metacognition. Implications for research and practice]* (pp. 77–95). Münster: Waxmann.
- von Stechow, E. (2006). PISA und die Folgen für schwache Schülerinnen und Schüler [PISA and its impact on weak students]. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 75, 285–292.
- Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C. (2011). Development of competencies across the life span. In H. P. Blossfeld, H. G. Roßbach, & J. 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.
- Wocken, H. (2000). Leistung, Intelligenz und Sozillage von Schülern mit Lernbehinderungen: Vergleichende Untersuchungen an Förderschulen in Hamburg [Performance, intelligence, and the social situation of students with learning disabilities: comparative analyses of special schools in Hamburg]. *Zeitschrift für Heilpädagogik*, 51, 492–503.
- Wocken, H., & Gröhlich, C. (2009). Kompetenzen von Schülerinnen und Schülern an Hamburger Förderschulen [Competencies of students at special schools in Hamburg]. In W. Bos & M. Bosen (Eds.), *KESS 7: Kompetenzen und Einstellungen von Schülerinnen und Schülern der Jahrgangsstufe 7* (pp. 133–142). Münster: Waxmann.
- Ysseldyke, J. E., Thurlow, M. L., Langenfeld, K. L., Nelson, R. J., Teelucksingh, E., & Seyfarth, A. (1998). *Educational results for students with disabilities: What do the data tell us? (Technical Report 23)*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Zimmermann, S., Gehrler, K., Artelt, C., & Weinert, S. (2012). *The assessment of reading speed in Grade 5 and Grade 9*. Bamberg: Otto-Friedrich-University, Nationales Bildungspanel.