

Händel, Marion; Artelt, Cordula; Weinert, Sabine

Assessing metacognitive knowledge: development and evaluation of a test instrument

Journal for educational research online 5 (2013) 2, S. 162-188

urn:nbn:de:0111-opus-84297



in Kooperation mit / in cooperation with:



<http://www.waxmann.com>

Nutzungsbedingungen / conditions of use

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

We grant a non-exclusive, non-transferable, individual and limited right to using this document.
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.
By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Mitglied der Leibniz-Gemeinschaft
Informationszentrum (IZ) Bildung
Schloßstr. 29, D-60486 Frankfurt am Main
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Marion Händel, Cordula Artelt & Sabine Weinert

Assessing metacognitive knowledge: Development and evaluation of a test instrument

Abstract

Educational processes in modern information societies require personal initiative – not only in institutional contexts but also in out-of-school and working environments. Besides cognitive competencies (e.g., reading or mathematical literacy), a person's ability to regulate his/her own learning processes is therefore of particular importance for successful learning and working. Metacognition is a central component in the process of self-regulated learning. It is defined as cognition about cognition and encompasses two components: the knowledge component and the regulation component. To better understand how metacognition evolves and how its development is influenced by cognitive and motivational components, the German National Educational Panel Study (NEPS) aims at assessing different aspects of metacognition as well as self-regulation in the respective phases of the lifespan. The present paper gives an overview on the general approach of assessing metacognition and self-regulation within the NEPS, focusing on the assessment of one specific component of metacognition, namely, metacognitive knowledge. Results from two pilot studies on a newly developed test instrument (with two experimentally varied versions of the test) for secondary school students are presented. The results concerning the metacognitive knowledge test are discussed not only with regard to further developments within the NEPS but also with regard to their theoretical and practical implications.

Keywords

Metacognition; Self-regulation; Competence assessment

Dr. Marion Händel (corresponding author), Department of Educational Psychology, University of Erlangen-Nuremberg, Regensburger Straße 160, 90478 Nuremberg, Germany
e-mail: marion.haendel@fau.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

Erfassung metakognitiven Wissens: Entwicklung und Evaluation eines Testinstruments

Zusammenfassung

Steigende Anforderungen der Informationsgesellschaft bedürfen einer eigenständigen und kontinuierlichen Aneignung von Wissen. Dies zeigt sich in schulischen Aus- und Weiterbildungskontexten ebenso wie in Beruf und Freizeit. Erfolgreiches Lernen erfordert daher neben kognitiven Kompetenzen (z.B. Lesekompetenz oder mathematische Kompetenz) insbesondere auch die Befähigung, das eigene Lernen zu regulieren. Metakognition ist eine zentrale Komponente im Prozess des selbstregulierten Lernens. Konzeptuell sind zwei Facetten der Metakognition zu unterscheiden, eine Wissens- und eine Regulationskomponente. Um Entwicklungsverläufe zur Metakognition und den Einfluss kognitiver und motivationaler Variablen abbilden zu können, bedarf es Längsschnittstudien. Ein Ziel des Nationalen Bildungspanels (National Educational Panel Study, NEPS) ist es daher, neben kognitiven Komponenten auch Metakognition bzw. Aspekte der Selbstregulation in den jeweiligen Phasen des Lebenslaufs zu erfassen. Nach einem Überblick über den Ansatz zur Erfassung des selbstregulierten Lernens (Metakognition bzw. Selbstregulation) im NEPS fokussiert der vorliegende Beitrag auf die Diagnose metakognitiven Wissens mittels eines neu entwickelten Instruments (z.T. in empirisch variierten Versionen) für Schülerinnen und Schüler der Sekundarstufe. Ergebnisse zweier Pilotstudien werden präsentiert und in Hinblick auf ihre theoretische und praktische Bedeutsamkeit, auch bezüglich weiterer NEPS-Entwicklungen diskutiert.

Schlagworte

Metakognition; Selbstregulation; Kompetenzmessung

1. Indicators of self-regulated learning within the German National Educational Panel Study (NEPS)

The ability to regulate learning processes in order to reach personal goals is a prerequisite for being successful in our rapidly changing information society. People need to self-regulate their learning and to make strategic decisions about what to learn. This concerns learning in institutional contexts, such as schools or universities, as well as behavior in out-of-school contexts and working environments (see Baumert, Bos, & Lehmann, 2000).

Self-regulated learning is a complex construct formed by different research traditions dealing with different terms and labels, which denote the same or at least strongly overlapping construct/s (see Dinsmore, Alexander, & Loughlin, 2008). In addition, the relationship of metacognition and self-regulated learning is also treated differently (see Veenman, van Hout-Wolters, & Afflerbach, 2006 for an over-

view). In line with several models (e.g., Boekaerts, 1997) we treat self-regulated learning as composed of motivational, metacognitive, and cognitive aspects interacting with each other.

The importance of metacognition and self-regulated learning is shown by its relationship to measures of ability, achievement, and/or competencies. Several studies show evidence that students who self-regulate their learning (Artelt, Naumann, & Schneider, 2010; Dresel & Haugwitz, 2008) or have high metacognitive competencies (Artelt et al., 2010; Thillmann, 2008; van Kraayenoord & Schneider, 1999) perform better than their counterparts with worse self-regulatory learning behavior or less metacognitive knowledge. Furthermore, students from different German school tracks¹ differ in their metacognitive competencies (cf. Artelt, Demmrich, & Baumert, 2001).

This paper focuses on the investigation of self-regulated learning and metacognition from a longitudinal research perspective accomplished in the German National Educational Panel Study – NEPS (see the editorial of this issue by Artelt, Weinert, & Carstensen, 2013). Due to the changing importance of metacognitive and self-regulative aspects during the lifespan, measurements are chosen with reference to lifelong learning and the demands imposed by this on the learner. Thus, indicators may vary in terms of content as well as in terms of their relative weight or their importance for a specific educational stage (cf. Weinert et al., 2011). For example, in kindergarten basal self-regulation is assessed by a measure of delay of gratification (Berendes et al., 2011). By contrast, for university students and adults, longer-term goal orientations, social and emotional goals, and aspects of self-regulation, such as resource or time management and self-motivation, take on significance (cf. Caprara et al., 2008; Schwinger, von der Laden, & Spinath, 2007; Stumm, Thomas, & Dormann, 2010). Therefore, they are assessed by self-report measures in these age cohorts (despite their flaws, see expert report by Roßnagel, Bittner, & Staudinger, 2009; see also Wohlkinger, Ditton, von Maurice, Haugwitz, & Blossfeld, 2011).

However, these self-regulatory aspects are not discussed any further within this paper as its focus is on metacognition and the corresponding assessment procedures.

1.1 Metacognition

Although several theoretical models of metacognition have evolved over the last decades, most models have in common that they separate three or at least two components of metacognition. According to several authors metacognition is divided into *knowledge about* and *control over/regulation of* one's own cognitive system (see Brown, 1987; Flavell, 1979; Schraw & Moshman, 1995). Some authors fur-

1 In most German Federal States school tracking is accomplished by way of ability tracking after Grade 4.

ther distinguish the regulation component into an experience (see Flavell, 1979) and a skill (Brown, 1987) component. Pintrich, Wolters, and Baxter (2000) use a three-tier model of metacognition and self-regulated learning, namely, metacognitive knowledge, metacognitive judgments and monitoring (equatable to metacognitive experiences), and self-regulation and control (equatable to metacognitive skills). The latter are also known as the procedural aspect of metacognition as opposed to metacognitive knowledge as the declarative aspect. However, as Ertmer and Newby (1996) argue, the knowledge and the regulation component of metacognition are not entirely separate aspects but rather interact with each other. In a current model Efklides (2011) describes a complex interplay between a task level, a person level (where metacognitive knowledge and skills are located), and an interaction level (where metacognitive experiences take place).

1.1.1 Metacognitive knowledge

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. Thus, it includes knowledge about the strengths and weaknesses of one's own memory and learning, about cognitive requirements of tasks (i.e., their complexity and difficulty) as well as knowledge about ways and means (e.g., general and domain-specific strategies) of attaining cognitive learning and achievement goals. Metacognitive knowledge about strategies means knowledge about effective methods of learning. According to Paris, Lipson, and Wixson (1983) the knowledge about different kinds of strategies can in turn be divided into declarative, procedural, and conditional strategy knowledge.

Declarative strategy knowledge is the awareness of strategies, that is, the awareness that a certain strategy exists. Procedural knowledge describes how a strategy works effectively, and conditional knowledge helps to understand which strategies are useful for solving a certain task. The development of metacognitive knowledge starts in kindergarten and continues to develop beyond adolescence over the entire lifespan (see Alexander & Schwanenflugel, 1996; Artelt, Neuenhaus, Lingel, & Schneider, 2012; Baker, 1989; Hasselhorn, 2006; Schneider & Lockl, 2006) as long as educational processes continue to challenge the learner (Veenman et al., 2006). According to Borkowski, Chan, and Muthukrishna (2000) metacognitive knowledge may develop from domain- and situation-specific knowledge (i.e., knowledge acquired in and intertwined with a specific content domain) to domain-general knowledge (i.e., knowledge transcending across domains). Recent research results involving fifth graders indicate – at least for this age group – that the construct of metacognitive knowledge is better represented by a multidimensional model with separate dimensions for specific (mathematics, reading comprehension) and general aspects of metacognitive knowledge than by a unidimensional model (Neuenhaus, Artelt, Lingel, & Schneider, 2011).

1.1.2 Metacognitive control and regulation

Metacognitive control and regulation is welded to concrete learning situations. As was already shown by Brown (1987) spontaneous control of one's own learning process may occur without the affected person being aware of it. Thus, it is not necessarily mirrored in explicit knowledge that can be verbalized. However, related perceptions and appraisals (e.g., judgments of learning, feeling of knowing, confidence judgments; see Nelson & Narens, 1990) which may occur before, during or after a problem-solving activity (Flavell, 1971, 1979) can be seen as prerequisites for an adequate use of self-regulation strategies and – at the same time – they are necessary within the process of acquiring metacognitive knowledge. The actual and conscious regulation of the learning process however takes place through planning, monitoring, and metastrategic activities, that is, the implementation of metacognitive knowledge in the process of self-regulated learning (see Flavell, 1979; Schneider & Artelt, 2010).

For metacognitive control and regulation no major developmental trend has been identified so far (cf. Lockl & Schneider, 2002; Schneider, 2008). It seems that these metacognitive skills initially develop in separate domains and, later on, become generalized across domains (Veenman & Spaans, 2005). However, this has not been empirically confirmed yet (Hasselhorn, 2006; Veenman et al., 2006) and, therefore, needs to be investigated by longitudinal studies.

1.2 Assessment procedures for metacognition

The assessment of different facets of metacognition requires different measurement procedures (for a substantial discussion of several assessment procedures, see Pintrich et al., 2000). Metacognitive knowledge is often assessed by so-called metacognitive awareness inventories with questionnaire character or by metacognitive knowledge tests. Control and regulation of cognition can be assessed by metacognitive judgments (experience component; referring to Efklides, 2011) or by the self-reported or actual strategy use (skill component; see Veenmann & Elshout, 1999).

Ideally, measures of metacognition inherit the feature of interpretability of the resulting indicators relative to a knowledge or performance standard (see Artelt & Neuenhaus, 2010; Artelt & Schneider, in press). Below, several assessment procedures are discussed with reference to the compliance with this criterion.

1.2.1 Assessing metacognitive knowledge

A well-established questionnaire for the purpose of measuring declarative, procedural, and conditional metacognitive knowledge is the Metacognitive Awareness Inventory by Schraw and Dennison (1994). However, like other self-reports, this

instrument as well as other questionnaires lack a clear benchmark of evaluation, and it is not clear whether the items are actually suitable to assess knowledge instead of usual behavior (see also Artelt & Neuenhaus, 2010; Artelt & Schneider, in press, or Maag Merki, Ramseier, & Karlen, in press for a discussion of the caveats of such measures). Currently developed instruments are also affected by this limitation (see the domain-specific instrument Metacognitive Awareness of Reading Strategies Inventory by Mokhtari & Reichard, 2002, or the metacognitive knowledge test about self, task, and strategies in mathematics developed by Vlachopoulos & Efklides, 2012). Another approach by Kreutzer, Leonard, and Flavell (1975) or Myers and Paris (1978) are interviews or open-ended questions to assess the awareness of critical variables influencing memory, learning, or domain-specific activities (see also the instruments described by Swanson, 1990, or van Velzen, 2012). Although providing possibly more detailed insights into metacognitive knowledge, such assessment methods are restricted to individual testing situations. Scenario-based tests, such as the Index of Reading Awareness by Jacobs and Paris (1987) or the Metacomprehension Strategy Index by Schmitt (1990) have been developed to assess the quality of metacognitive knowledge against an evaluation standard in the domain of reading (i.e., conditional knowledge about an effective use of strategies). Such tests consist of several scenarios describing typical challenging situations; each scenario is followed by a list of approaches of differing strategic quality (multiple-choice tests, single selection).

Several other scenario-based tests have been developed over the last decade (e.g., the metacognitive strategy knowledge test concerning reading strategies implemented in several languages within the 34 OECD-states participating in the Programme for International Student Assessment (PISA), see Artelt, Beinicke, Schlagmüller, & Schneider, 2009). In the following we describe those tests in more detail as they build the basis for the development of a NEPS test. Those tests contain multiple-select items that are scored via so-called pair comparisons (option X is more or less useful than option Y) with reference to experts' judgments of the relative usefulness of the presented alternatives. It was shown that these test instruments are reliable (e.g., Cronbach's $\alpha = .84$ with the PISA test) and economical in use, that they refer to concrete learning situations, and that they are interpretable against a clear benchmark (Artelt & Schneider, in press). The tests ask for metacognitive knowledge in different content domains (see Schlagmüller & Schneider, 2007, or Artelt et al., 2009 for tests on metacognitive reading strategy knowledge; Marschner, 2011, Scherer & Tiemann, 2012, or Thillmann, 2008, for tests on metacognitive experimental strategy knowledge; Neuenhaus et al., 2011 for tests on domain-specific and domain-general metacognitive knowledge and Dubowy, 2010, or Ebert, 2011, for tests on metamemory). The tests are developed for the group of kindergarten children (Dubowy, 2010, or Ebert, 2011) or secondary school students (the reading strategy knowledge test for Grade 7 to 12 (WLST) by Schlagmüller & Schneider, 2007 for students up to Grade 12 but with ceiling effects from Grade 9 on). Tests for younger children usually refer to memory and learning

in general (domain-general), whereas the others refer to a single domain (domain-specific), mostly the domain of reading (e.g., WLST).

Whereas the domain-specific tests focus on particular requirements relevant only to a certain content domain, the general metacognitive knowledge tests consist of tasks relevant to several situations that are not restricted to a particular subject or content domain. However, as mentioned by Neuenhaus et al. (2011), the latter still have to be considered as specific with respect to the situations they refer to. The scenarios for older students are mostly borrowed from school-relevant activities (e.g., exam preparation), whereas the scenarios for younger students are mostly related to leisure time (e.g., remembering personal tasks). The format of the answer scale varies between direct pair comparisons (forced choice) and six rating points (e.g., 6-point rating scale of usefulness within the PISA instrument, Artelt et al., 2010). However, the scoring via pair comparisons oftentimes requires only a two-stage differentiation between more and less useful strategies (Artelt et al., 2009). The number of given alternate options to master the presented learning problem usually varies from three up to seven. A higher number of alternatives imply (a) higher demands on reading comprehension due to more textual information and (b) increased difficulty as it should be easier to compare only two or three alternatives than to rate seven alternatives according to their usefulness in the respective situation. Finally, the person addressed in the test situation differs between and even within various test instruments: The perspective varies between the second-person perspective (imagine you did ...) and the third-person perspective (he/she does ...). Also, the personal address of given options varies between first-person perspective (I would do) and third-person perspective (he/she does/should do).

1.2.2 Assessing prerequisites for metacognitive control and regulation

Prerequisites for metacognitive monitoring can be measured by assessing an individual's metacognitive judgment about his or her learning and performance, that is, via confidence judgments or judgments of performance accuracy (see Boekaerts & Rozendaal, 2010; Hines, Touron, & Hertzog, 2009; Nelson & Narens, 1990; Schraw, 2009). These judgments refer to a concrete learning situation and can be inquired before, during, or after learning or test processing. These measurements allow for a comparison with actual achievement and therefore maintain a clear evaluation standard.

Other ways of assessing metacognitive control and regulation make use of either online (i.e., during the learning activity) or offline (by way of self-reports after the learning activity) measures; cf. Wirth and Leutner (2008). Self-report measures, such as questionnaires assessing strategy use retrospectively, are related to more or less broadly defined learning occasions in the past (i.e., LASSI by Weinstein & Palmer, 1990; MSLQ by Pintrich, Smith, Garcia, & McKeachie, 1991).

They have been criticized because they often miss the possibility to evaluate the quality and appropriateness of strategy application according to a concrete learning situation. Further criticism also refers to their high demands on memory and the low correlations between the resulting strategy scores and student achievement (Artelt, 2000; Spörer & Brunstein, 2006; Veenman et al., 2006). Yet other methods that also provide a clear benchmark are, for example, measures for error detection or allocation of study time (Dubowy, 2010; Lockl & Schneider, 2003; Metcalfe, 2009; Pieschl, 2009), thinking-aloud protocols (Veenman, 2005) or log-file-based analyses for computer-based learning (Azevedo, 2007); for an overview see, for example, Zimmerman (2008).

1.3 Assessment of metacognition within the NEPS

Indicators of metacognition within the NEPS are chosen in accordance with the relevance of specific metacognitive aspects throughout the life course (Weinert et al., 2011). As research on metacognition uses different instruments and approaches that vary – among other things – according to the grain size of the measures (Veenman et al., 2006), the NEPS implements measures tapping the same constructs and following the same construction principle and grain size for as long as possible. Both the metacognitive knowledge and the metacognitive control are part of the assessment strategy within the NEPS.

1.3.1 Assessing metacognitive knowledge within the NEPS

For the NEPS student cohorts metacognitive knowledge is coherently assessed by a scenario-based metacognitive knowledge test. As mentioned before, the scenario-based knowledge approach has already been administered in different PISA cycles (Artelt et al., 2001, 2010). It was chosen because of its advantages such as testing within a group setting or because of its clear benchmark criterion. The test primarily focuses on different aspects of strategy knowledge. In order to obtain valid indicators of metacognitive knowledge within the NEPS and because previously developed instruments differ in a number of features, new instruments have had to be developed that are suitable for the special demands of the longitudinal approach. The decision for the implementation of the specific features used in the NEPS test was based on theoretical considerations as well as on the empirical studies presented below.

1.3.2 Assessing prerequisites of metacognitive control and regulation within the NEPS

Metacognitive indicators of metacognitive control and regulation within the NEPS are assessed throughout the whole lifespan (except in the very early stages). Due to practical constraints we do not directly assess metacognition through activities of planning, monitoring, or regulation but as judgments of performance as indicators for metacognitive control and regulation. The assessment is integrated in each competence domain of the NEPS study (cf. Weinert et al., 2011) by single indicators of judgments of performance directly after the test phase (retrospective judgment/post-diction, see Nelson & Narens, 1990; Schraw, 2009). Subjects are asked to estimate their own achievement in the respective domain-specific competence test by judging the number of correctly given answers (retrospective judgment of performance accuracy according to Schraw, 2009; see also Maki, Shields, Wheeler, & Zacchilli, 2005). Thus, NEPS data will give substantial insight into the development of metacognitive monitoring in different content domains.

1.4 Purpose and research questions

Current instruments on metacognitive knowledge differ in several content and technical components. To our knowledge, the different aspects and their influence on test validity have not been empirically examined so far (however, Pintrich et al., 2000 already discussed several characteristics of the Index of Reading Awareness and the Metacognitive Awareness Inventory in comparison to each other). We assume that different test construction criteria impact on the validity of the tests result. Therefore, we are paying attention to distinct criteria in all scenarios with the aim of developing an economical, reliable, and valid scenario-based metacognitive knowledge test for the assessment in the NEPS. With reference to the NEPS as a longitudinal study, we aim at assessing *domain-general* metacognitive knowledge detached from particular school subjects or content domains, because these situations lose their practical relevance after schooling. Furthermore, due to practical reasons (e.g., limited assessment time) it is not possible to assess metacognitive knowledge for several content domains separately.

We evaluate the newly developed test with respect to the following issues. Our first attempt is to investigate whether the perspective of the items influences the assessment. We hypothesize that test scores are lower in a first-/second-person perspective than in a third-person perspective. In the first-/second-person perspective students might rate the options in the sense of what *they would do* in a given situation instead of rating the *objective appropriateness and usefulness* of the options available. That is, a third-person perspective is ought to measure knowledge more validly than a first-/second-person perspective, which probably measures the quality of typical behavior. Hence, scores should be lower in the latter version, because students do not always implement the most useful strategies in spite of

knowing them and being aware of them (production deficit, see Hasselhorn, 2006), because students might not always be motivated to use those strategies, or because they might think that they are smart enough already and, therefore, do not need to implement any strategies. In addition, we assume that the score of a third-person perspective version should correlate with self-reported frequency of strategy use to a lesser degree than the score of a first-/second-person perspective version. Moreover, we are interested in the appropriateness of test difficulty for students of different grades. To investigate the psychometric properties of the correspondingly constructed NEPS test, we check whether we have succeeded in the construction of a unidimensional test for general (in contrast to subject-specific) metacognitive knowledge for students of Grade 6 and 9 of all regular school types in Germany, and test for reliability (internal consistency) and validity as quality criteria. Content validity is examined by expert ratings and correlations are analyzed with school achievement in mathematics and in German lessons (the lingua franca) as well as with reading competence (in German). Furthermore, we test for gender and school type differences. On the basis of findings from research literature (i.e., Artelt et al., 2010) we assume that girls should outperform boys on the metacognitive knowledge test. Given that metacognitive knowledge increases if learners are exposed to challenging and relevant educational processes and in consideration of findings that high-performing students tend to have a richer knowledge base for metacognitive knowledge, we assume that school type differences (cf. Artelt et al., 2001) do occur.

2. Method

2.1 Sample and general procedure

Two pilot studies with students from secondary school were conducted to evaluate the newly developed metacognitive knowledge tests. In both studies the tests were administered in regular classrooms, and the students received an incentive for their participation. In the first pilot study we implemented an experimental design with varying treatments, whereas the second pilot study aimed at testing the psychometric properties of the final test. The test was scored with reference to ratings provided by experts. Following this procedure, we established content validity for the developed test.

2.1.1 Pilot Study 1

In the first pilot study, 212 female students of a Bavarian secondary school (*Gymnasium*) participated. Students were sampled from Grade 5 to 9 and their mean age was 13.4 years ($SD = 1.6$). By collecting data from students in Grade 5 to Grade 9 we were able to analyze age differences concerning metacognitive knowl-

edge. The main aim of the first study was to investigate the influence of different test versions. In the experimental study, students were given 5 minutes to fill in a questionnaire about their usual learning behavior. Afterwards students were asked to answer the metacognitive knowledge test within 20 minutes. Two content-identical versions were realized, differing only in the perspective (see Section 2.2 for a detailed description). Participating students were randomly assigned to one of two test versions within each participating class (109 students with the first-/second-person perspective and 103 students with the third-person perspective). The two groups did not differ in relevant variables (see below). Table 1 shows the sample statistics separately for the different grades.

Table 1: Sample statistics of Study 1

School grade	N	Age [years]
	Total number (percentage)	M (SD)
Grade 5	49 (23.1)	11.5 (0.4)
Grade 6	38 (17.9)	12.5 (0.5)
Grade 7	43 (20.3)	13.4 (0.5)
Grade 8	45 (21.2)	14.5 (0.4)
Grade 9	37 (17.5)	15.7 (0.5)

2.1.2 Pilot Study 2

In the second pilot study, sixth-grade students from four Federal States of all school types participated. The study was a NEPS longitudinal pilot study conducted by the IEA Data Processing and Research Center (see Blossfeld, von Maurice, & Schneider, 2011, for an overview of the NEPS design). Altogether, 366 students of 21 schools participated (52.7% female) with a mean age of 12.0 years ($SD = 0.7$). The sample statistics of the second pilot study are displayed in Table 2. The proportion of students attending the different school types is representative for German schools (cf. Statistisches Bundesamt, 2011). Hence, the percentage of students per school type differs because more students attend the school type *Gymnasium* than, for example, the school type *Hauptschule*. Also, the proportion of female students differs between the school types. A subsample of 172 students participated in a study conducted one year earlier.

Table 2: Sample statistics of Study 2

Type of school ^a	<i>N</i> Total number (percentage)	% Female	Age [years] <i>M</i> (<i>SD</i>)
Hauptschule	37 (10.1)	40.5	12.1 (0.8)
Schule mit mehreren Bildungsgängen ^b	53 (14.5)	54.7	12.1 (0.5)
Realschule	120 (32.8)	48.3	12.1 (1.0)
Gesamtschule	41 (11.2)	65.9	12.1 (0.5)
Gymnasium	115 (31.4)	55.7	11.8 (0.4)

^aThe school types are sorted in the order of the degree students can achieve at a specific school type.

^bSchools that combine the courses of education provided at Hauptschule and Realschule.

Several competence tests (scientific literacy, metacognitive knowledge, and listening comprehension; in this order) and questionnaires were implemented in this longitudinal pilot study. For the purpose of this paper, we focus on the relevant variables to describe the metacognitive knowledge test, including data of a reading competence test collected one year earlier. The participating students were again given 20 minutes to answer the test items of the metacognitive knowledge test.

2.2 Measurements

In both pilot studies, the metacognitive knowledge test and instruments assessing several control variables were administered.

To measure *metacognitive knowledge* a total item pool of 15 scenarios was constructed on the basis of previous work on metacognitive knowledge tests (see above). In the process of test construction the scenarios were modified and optimized through qualitative cognitive interviews with students as well as teachers and as a result of discussions in an expert team. We selected eight scenarios for the final NEPS metacognitive knowledge test scheduled to take 15 minutes. The descriptive statistics of the test implemented in the main study are displayed in Lockl (2012). Seven of those scenarios were implemented in each of the pilot studies (one additional scenario was piloted in Study 2 only). The item selection process was based on different criteria: Item analyses were conducted to choose items with high discriminatory power and adequate item difficulty. In addition, we attempted to achieve a broad range of possible strategies and contents presented in the scenarios. We furthermore took into account the comments made by the experts with regard to the proposed strategies and scenarios. The chosen scenarios are supposed to be concrete, specific, and age appropriate.

To interpret and describe the results of the two studies more easily, only the seven scenarios presented in both studies are included in the following analyses (besides, results concerning the experimental variation of the test version in Pilot

Study 1 do not differ from the total item pool in this study). The scenarios include various application contexts ranging from learning from texts and text comprehension to leisure time activities. In particular, four 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. Cognitive, metacognitive, and resource management strategies were considered resulting in a knowledge test about solving cognitive tasks, such as remembering or organizing information, but also about planning and regulating learning as well as leisure time activities, and about general learning requirements, such as using resource management strategies. Each task consists of two parts, a scenario with a relevant learning situation and a specific cognitive demand (e.g., “remember a list of items”) and six strategies that vary in their degree of effectiveness for the given situation. Cognitive interviews showed that students were able to rate six statements.

By way of example, we describe in the following a domain-general school-related learning scenario as well as a scenario concerning leisure-time activities with three corresponding sample statements each. One of the domain-general school-related scenarios deals with the self-evaluation in the process of exam preparation. Sample statements are concerned with solving items of an exercise book, discussing relevant topics with peers or persuading oneself that one will be successful. One out-of-school scenario is concerned with time and resource management for leisure-time activities, namely, for organizing a birthday party. Proposed statements are, for example, to create a list and tick off completed tasks, to permanently think about the exciting event, or to organize everything just on the day before the party. The type and structure of items can be seen in Figure A1 in the appendix (see also Lockl, 2012).²

The number of rating points on the scale of usefulness was set to 4, because we do not expect a more fine-grained differentiation of the alternatives (cf. Artelt et al., 2010). However, implementing only a 2-point scale would restrict the answer behavior (loss of variance), raise the probability of guessing, and simplify the test too much. In addition, asking directly for pair comparison using a forced-choice format might irritate and discourage students due to excessive reading effort. In sum, the reported scenarios are characterized by the following features:

- Contexts: school and leisure time activities, seven scenarios altogether;
- Strategies: cognitive, metacognitive, and resource management;
- Number of presented strategies for each scenario: six (42 strategies altogether);
- Rating scale: 4-point scale of usefulness (1 = *not at all useful*, 4 = *very useful*).

For Pilot Study 1, we developed two different test versions that differ only in the phrasing of the personal perspective. In Version A, the scenarios are phrased in the second-person perspective and given alternatives in the first-person perspective. In

2 Due to the longitudinal design of the study it is not possible to provide an original sample item of the final item pool.

Version B, which represents the target version for the NEPS and which was realized in Pilot Study 2, we consistently used the third-person perspective.

In addition to the metacognitive knowledge test the following variables were collected. *School achievement* data of the participating students were collected in both pilot studies via school grades in the subjects mathematics and German (the lingua franca). School grades (1–6) were recoded, so that a high grade resembles a good performance. In addition to these indicators, a learning questionnaire about the frequency of strategy use was implemented in Pilot Study 1 to investigate the relation between metacognitive knowledge (tested with the two versions differing in perspective) and self-reported strategy use. The *learning questionnaire* about the frequency of strategy use comprises 20 items on a 4-point Likert scale. The items were borrowed from the KSI (Kieler Lernstrategien-Inventar) by Baumert, Heyn, and Köller (1992). Each four items were administered to assess the strategy use of surface strategies (rehearsal strategies, Cronbach's $\alpha = .61$) and of deep-level strategies (elaboration strategies, Cronbach's $\alpha = .73$). The subscale metacognitive strategies encompasses 12 items about planning, monitoring, and regulating strategy use (Cronbach's $\alpha = .71$). *Reading competence* in Pilot Study 2 was assessed by the NEPS reading competence test (see Gehrler, Zimmermann, Artelt, & Weinert, 2013, this issue).

3. Results

3.1 Expert rating

Based on theoretical assumptions about the adequacy of strategies for a specific situation or task the test includes 84 pair comparisons. A pair comparison is considered to be valid for the assessment of metacognitive knowledge if experts agreed to at least 75% in the direction of the pair comparison (i.e., 75% or more of the experts rated a strategy alternative as superior or subordinate to another). According to the ratings of 10 experts, 63 of the 84 theoretically assumed pair comparisons reached this criterion (at least 75% agreement). The number of valid pair comparisons for a scenario ranged between 7 and 11 comparisons.

3.2 Results of Pilot Study 1

The main purpose of Pilot Study 1 was to investigate the effect of the phrasing of two different test versions. Firstly, we inspect whether the students who were randomly assigned to one of the two conditions differed in the collected variables (multivariate variance analyses of age, school grades, and the self-reported strategy use as dependent variables and version of the metacognitive knowledge test as independent variable). As expected, students of the two treatments did not differ

with respect to relevant background variables (i.e., grades and age) or the frequency of self-reported strategy use ($F(6, 201) = 0.52, p = .79$).

Treatment differences in metacognitive knowledge were tested by an independent *t* test. As expected, students who filled in the first-/second-person perspective test scored significantly lower than students who worked on the third-person perspective test ($t(210) = 2.91, p < .005, d = 0.46$; see Table 3 for the descriptive statistics).

Table 3: Descriptive statistics of the metacognitive knowledge test score (%-score) by test version

Item perspective	<i>N</i>	Test score	
		<i>M (SE)</i>	<i>SD</i>
First-/second-person	109	64.2 (1.4)	14.3
Third-person	103	69.6 (1.2)	12.6

Furthermore, we analyzed the correlations of the test scores with the self-reported strategy use for the two versions separately. As expected, the correlation coefficient (Pearson's *r*) of the self-reported metacognitive strategy use with the metacognitive knowledge seems to be higher for the first-/second-person perspective ($r = .25, p < .05$) than for the third-person perspective ($r = .13, p = .18$). However, testing for statistical differences between the correlations by a test of difference between the *z*-standardized correlation coefficients shows no significant difference (raw difference = 0.12, $z = .76, p = .39$). In addition, no significant differences between correlations could be detected for the cognitive learning strategies (rehearsal and elaboration).

Finally, we investigated grade differences for the subsample that filled in the third-person perspective test version. A univariate analysis of variance with metacognitive knowledge as dependent variable and grades as independent variable shows that test scores differed between the grades ($F(4, 98) = 6.26, p < .001, \eta^2 = .20$) with older students scoring higher than younger students (see Table 4).

Table 4: Descriptive statistics of the third-person perspective metacognitive knowledge test score (%-score) for students of different grades

Grade	N	Test score	
		M (SE)	SD
5	23	62.6 (2.5)	11.8
6	17	62.7 (3.4)	14.0
7	21	72.9 (2.2)	10.2
8	23	76.1 (1.9)	9.3
9	19	72.6 (2.8)	12.3
all	103	69.6 (1.2)	12.6

3.3 Results of Pilot Study 2

Investigating the psychometric properties of the newly developed instrument, the following results were obtained: The mean test score is 62.6% ($SD = 16.7$) for the investigated sample. The mean scores of the seven single scenarios differed from $M = 51.4\%$ ($SD = 26.8$) to $M = 69.3\%$ ($SD = 28.7$). The internal consistency for the seven scenarios of the metacognitive knowledge test is satisfying (Cronbach's $\alpha = .76$), which indicates that we were successful in constructing a homogeneous test. To investigate whether the test is in fact unidimensional, we compared a uni-versus two 2-dimensional models with school versus out-of-school and domain-specific versus domain-general scenarios loading on separate factors. To this end, we conducted three confirmatory factor analyses (CFAs) by *Mplus* Version 6.11 (Muthén & Muthén, 1998–2011) and compared the fit statistics of the nested models (see Table 5 and Table 6 for the results of the CFA). The coefficients for all models indicate a good fit with a slight advantage of the two 2-factor solutions. However, the change of χ^2 is not significant (χ^2 -change = 1.56 or 1.43, respectively, which is both smaller than the threshold of 3.84), showing that there is no clear advantage of a two-dimensional model as compared to the unidimensional one.

Table 5: Fit indices for the three proposed models

Fit values	1-factor solution	2-factor solution	
		School vs. out-of-school	Domain-specific vs. domain-general
χ^2 -fit (336)	21.15, $p = .10$	19.59, $p = .11$	19.72, $p = .10$
RMSEA	.039	.039	.039
CFI	.983	.984	.984

Note. Model 1: one-factor solution with all scenarios loading on one factor; Model 2: two-factor solution with school- and out-of school items loading on two separate factors; Model 3: two-factor solution with domain-specific and domain-general scenarios loading on two separate factors.

Table 6: Unstandardized loadings and standardized loadings (standard errors) for the three proposed structural models

Number and characteristics of the scenarios	1-factor solution				2-factor solution: School vs. out-of-school				2-factor solution: Domain-specific vs. domain-general			
	Unstandardized		Standardized		Unstandardized		Standardized		Unstandardized		Standardized	
	λ	SE	λ	SE	λ	SE	λ	SE	λ	SE	λ	SE
1: Out-of-school/general	1.00	0.00	0.66	0.04	1.00	0.00	0.67	0.04	1.00	0.00	0.67	0.04
4: Out-of-school/general	1.16	0.13	0.65	0.04	1.17	0.13	0.66	0.04	1.17	0.13	0.65	0.04
7: Out-of-school/general	0.89	0.12	0.53	0.05	0.88	0.12	0.54	0.05	0.89	0.12	0.54	0.05
5: School/general	0.65	0.11	0.39	0.05	1.00	0.00	0.40	0.06	0.65	0.11	0.39	0.06
6: School/general	0.81	0.11	0.50	0.05	1.26	0.23	0.51	0.05	0.81	0.11	0.50	0.05
2: School/specific	0.95	0.11	0.61	0.04	1.47	0.26	0.62	0.05	1.00	0.00	0.65	0.05
3: School/specific	0.87	0.11	0.55	0.05	1.36	0.24	0.57	0.05	0.91	0.12	0.58	0.05

Note. The scenarios were sorted according to the sequence of scenarios in the test booklet, the domain-specificity/-generality, and the context of the scenarios (out-of school vs. school scenarios).

Moreover, we analyzed the effects of gender and school type as well as the relationship with other achievement variables. We conducted an analysis of variance with gender and school type as independent factors and metacognitive knowledge as dependent variable. The according descriptive statistics (means and standard deviations) are displayed in Table 7. The results show a significant gender difference ($F(1, 363) = 3.31, p < .05, \eta^2 = .01$) with girls scoring higher than boys. The effect of school type is significant ($F(4, 363) = 21.50, p < .001, \eta^2 = .20$) with students of the school type Gymnasium scoring higher than students of all other school types (Hochberg's GT2 was used as post-hoc procedure due to different cell sizes and indicates mean differences from 12.8% to 21.3%, each $p < .001$). No further comparisons reached statistical significance. No significant interaction effect of gender and grade could be detected ($F(4, 363) = 0.84, p = .50, \eta^2 = .01$).

Table 7: Means and standard deviations of the metacognitive knowledge test score (%-score) by school type and gender

School type ^a	Gender	N	Test score			
			Weighted		Unweighted	
			M	(SD)	M	(SD)
Hauptschule	Male	21	50.0	(15.0)	50.0	(15.0)
	Female	15	54.7	(14.4)	55.2	(17.4)
	Total	37	51.2	(15.3)	56.9	(18.4)
Schule mit mehreren Bildungsgängen ^b	Male	24	55.3	(17.4)	62.3	(16.5)
	Female	29	62.1	(16.6)	70.0	(14.1)
	Total	53	59.0	(17.1)	60.2	(17.7)
Realschule	Male	60	56.9	(18.4)	54.7	(14.4)
	Female	58	58.9	(15.0)	62.1	(16.6)
	Total	120	57.5	(17.0)	58.9	(15.0)
Gesamtschule	Male	14	62.3	(16.6)	59.5	(15.0)
	Female	27	59.5	(15.0)	75.8	(8.6)
	Total	41	60.4	(15.4)	64.7	(15.5)
Gymnasium	Male	51	70.0	(14.1)	51.9	(14.7)
	Female	64	78.8	(8.6)	59.0	(17.1)
	Total	115	73.2	(11.7)	57.9	(16.7)
Total	Male	170	60.0	(17.7)	60.5	(15.4)
	Female	193	64.7	(15.5)	73.2	(11.7)
	Total	366	62.3	(17.0)	62.6	(16.7)

^a The school types are sorted in the order of the degree students can achieve at a specific school type.

^b Schools that combine the courses of education provided at Hauptschule and Realschule.

To investigate the relationship of metacognitive knowledge and other achievement variables, we calculated bivariate correlations. We report Spearman's ρ for the correlations with school grades (as German grades are not considered interval data) and Pearson's r for the correlation with reading competence (data available for 172 students). For students of all school types, the metacognitive knowledge score significantly correlates with the school grade in the subjects German ($\rho = .29$, $p < .001$) and mathematics ($\rho = .26$, $p < .001$). The relationship of reading competence and metacognitive knowledge is significant ($r = .51$, $p < .001$).

4. Summary and discussion

Metacognition and self-regulation are considered key competencies in the 21st century. In institutional contexts as well as in out-of-school or working environments, metacognition and self-regulation are of particular importance for successful learn-

ing and working. The NEPS aims at assessing these components longitudinally over several educational stages of the life course.

Metacognitive knowledge and metacognitive control, the two components of metacognition, as well as other aspects of self-regulation are all part of the NEPS assessment strategy. In this paper we introduced the measurements used to assess indicators of self-regulated learning within the NEPS. The focus of the paper was on two pilot studies examining the development and evaluation of the metacognitive knowledge tests for secondary students in the NEPS. During the test development phase several cognitive interviews with students and teachers were conducted. The newly developed test instrument has several unique features that make it distinguishable from previously constructed tests, which focused mostly on domain-specific metacognitive knowledge and sometimes lack a clear evaluation standard: The test meets the requirements of the NEPS study with regard to available assessment time and administration in group settings and consists of several scenarios featuring challenging situations from school as well as from leisure-time contexts. The students' task was to rate the proposed strategy alternatives according to their usefulness in the given situation. In addition to these theoretically driven characteristics, technical components were fixed – such as the number of given approaches/strategies (six) and the number of rating points on the Likert scale (4-point scale). Finally, to assure content validity of our test and also to get a clear standard of evaluation, we obtained expert ratings. Nevertheless, the validity of the test is limited due to the contents implemented in the test that presumably do not cover the wide range of metacognition (cf. Pintrich et al., 2000).

To investigate our research questions we conducted two pilot studies. The two studies followed different aims and featured different samples (gender distribution, age, and school type). Although the results of the two pilot studies cannot be compared directly with one another they complement each other to some degree: The first pilot study gives some insights into age differences and resulted in the derivation of important construction principles for the NEPS test. Differences relating to the person's perspective of the situations and strategic alternatives were tested regarding test scores as well as their correlations with self-reported strategy use. Pilot Study 2 was built on the results of the first study and explored the structure and validity of the newly developed test. One particular shortcoming of the first pilot study was the biased sample size – only girls from different grades in the highest achieving school type participated in this study. In the second study, we were able to realize a more representative sample from Grade 6 of both genders – the target population of the NEPS test.³ The sample was recruited within a NEPS longitudinal pilot study aiming at producing similar sample composition as in the NEPS main studies. Finally, the NEPS main studies will give further evidence on the validity of the tests.

3 Metacognition data for Grade 9 were unfortunately not available as other competence tests had been piloted in the relevant pilot study.

The results have confirmed our assumptions that test scores are significantly higher in the third-person perspective version than in the first-/second-person perspective version. Although this result is restricted to female students (as only girls participated in the study) we assume that the perception of the two different ways of phrasing should not be gender-specific (or at least not be opposite for boys). We trace the result back to the link students make when they read the items in the first-/second-person perspective. When rating the given approaches/strategies in the first-/second-order perspective, students probably tend to rate what they would do instead of rating the objective usefulness of the strategies in the given situation. For several reasons personal behavior sometimes differs from the perfect strategic behavior. For the presented test providing alternate strategies we assume that especially the following two mechanisms apply: (a) People either judge tasks as too easy or overestimate their own skills and, therefore, try to solve them without effortful strategy use or (b) people are not motivated to use these strategies. This might lead to inadequate rating in the first-/second-person perspective.

In accordance with the PISA results self-reported strategy use correlated only slightly with the test scores of the metacognitive knowledge test. Although the correlations for the third-person perspective were numerically lower than in the first-/second-person perspective, the difference did not reach statistical significance. One explanation might be that test scores in both versions were relatively high (cf. Table 3), especially in the higher grades. In addition, the test version in the first-/second-person perspective is assumed to assess the quality of one's own behavior which is still different from the format and purpose of questionnaires. Varying the personal perspective of the scenarios within each participant could, in an additional study, provide more empirical evidence on how individuals differ in accordance with personal perspectives.

The relatively high test scores in Pilot Study 1 can be traced back to two reasons. First, only girls and only the academically higher achieving students from Gymnasium, who are supposed to have higher metacognitive knowledge, participated in the study. By contrast, the sample of the NEPS main study will be representative including students of types of schools, so that the test should be of adequate difficulty for this more heterogeneous sample. Second, the sample of this pilot study ranges from fifth to ninth graders. As expected older students reached higher test scores than younger students, which indicates that metacognitive knowledge develops during secondary school. As a consequence, we need to construct more difficult items to avoid ceiling effects with the older age group of students. A first pilot study with university students reconfirmed the challenge in constructing test items of adequate difficulty (see Händel, Tupac-Yupanqui, & Lockl, 2012). The NEPS main study as a longitudinal study will give further insights into the developmental processes.

All in all, the results underline the importance of item phrasing, especially the personal perspective in scenario-based metacognitive knowledge tests. Using a third-person perspective in the item phrasing of the metacognitive knowledge test further supports the assessment of metacognitive knowledge in contrast to the

quality of the respondent's own behavior. These results contribute not only to technical test-construction principles but furthermore have theoretical as well as practical implications: Also from an educational perspective there is a need to distinguish between *knowledge about* and *use of* learning strategies (see production deficit described, for example, by Hasselhorn, 2006, stating that learners may possess knowledge but not use it spontaneously). Only if researchers and educators know why their students choose not to use adequate learning strategies – because they are lacking metacognitive knowledge or motivation or due to an overestimation of their own skills – will they be able to apply a useful training method.

Results of Pilot Study 2 show that the internal consistency of the metacognitive knowledge test is satisfactory. In addition, the confirmatory factor analyses show good fit indices for the three theoretically driven factor analyses with no clear advantage of a two-dimensional model as compared to the unidimensional one. That is, neither the two-dimensional model with the factors school and out-of-school scenarios nor the two-dimensional model with the two factors domain-specific and domain-general scenarios showed better fit indices than the unidimensional model. Interpreting both two-factorial models, we have to be aware of the fact that out-of-school scenarios resemble domain-general scenarios and vice versa. Still, we can conclude that none of the theoretically driven two-factorial models outmatch the unidimensional model. However, our results are constrained by the number of constructed scenarios and do not provide information about the overall structure of metacognitive knowledge.

By comparing the results of preexisting groups who are supposed to differ in metacognitive knowledge we were able to provide information about the external validity of the test in the second pilot study (see Pintrich et al., 2000). The results show typical patterns: Girls score higher than boys (see Artelt et al., 2010) and school differences were shown to result in the expected direction. Due to the varying distribution of girls and boys in the different school types, we reported both weighted and unweighted means in Table 7, which displays that the direction remains identical. That is, students of Gymnasium score higher than students of all other school types (cf. Artelt et al., 2001).

Furthermore, the results are consistent with previously reported data, showing positive moderate correlations with grades in several school subjects. In addition, the metacognitive knowledge score correlates positively with reading competence and the correlation is of comparable size to the reported correlation for the PISA 2009 test on metacognitive knowledge and reading competence (see Artelt et al., 2010). As the reported results are restricted to Grade 6 students, data analyses with the NEPS main study in Grade 9 will reveal whether the results of our test can be replicated for older students (see Lockl, 2012). Further research considering the main study will allow for analyzing the interrelations with other domain-general and domain-specific competencies as well as with motivational variables and their reciprocal influences during development. In addition, further information about the external validity of the constructed test can be achieved by analyzing the correlations with students' cognitive abilities.

In sum, our pilot studies contribute significantly to research on the assessment of metacognitive knowledge, and they promote the assessment of metacognitive knowledge within the NEPS. We have succeeded in constructing and evaluating a metacognitive knowledge test with scenarios including different contexts and a broad range of strategies. Furthermore, the test has been constructed as an objective test with items formulated in the third-person perspective and with strategies that have to be rated according to their usefulness in the given situation. However, our results are limited to the cohort of secondary students. Future challenges will be to construct a test of adequate difficulty for the cohort of university students and adults, who are assumed to have richer knowledge. The work by Maag Merki et al. (in press) who already developed and evaluated a test for older students and university students may be a helpful guidance in the construction of a test instrument for this cohort. An important issue for the construction process is to find adequate scenarios embedded in contexts relevant to adults, but also to construct tests for young students with little reading experience. First results of the test development for first graders have shown how the established construction principles can be transferred to a test for younger students (see Lockl, Händel, Haberkorn, & Weinert, 2013).

Acknowledgments

We thank Kathrin Lockl and Nora Neuenhaus for their input and discussion on the assessment of metacognition as well as Juliane Fischer and Verena Waliczek for their support in the process of item development and for their engagement in the data collection for the pilot studies. Finally, we thank the experts for their ratings of the items.

References

- Alexander, J. M., & Schwanenflugel, P. J. (1996). Development of metacognitive concepts about thinking in gifted and nongifted children: Recent research. *Learning and Individual Differences, 8*, 305–325.
- Artelt, C. (2000). Wie prädiktiv sind retrospektive Selbstberichte über den Gebrauch von Lernstrategien für strategisches Lernen? *Zeitschrift für Pädagogische Psychologie, 14*, 72–84.
- Artelt, C., Beinicke, A., Schlagmüller, M., & Schneider, W. (2009). Diagnose von Strategiewissen beim Textverstehen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 41*, 96–103.
- Artelt, C., Demmrich, A., & Baumert, J. (2001). Selbstreguliertes Lernen. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann, & M. Weiß (Eds.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 271–298). Opladen, Germany: Leske+Budrich.
- Artelt, C., Naumann, J., & Schneider, W. (2010). Lesemotivation und Lernstrategien. 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. 73–112). Münster, Germany: Waxmann.
- Artelt, C., & Neuenhaus, N. (2010). Metakognition und Leistung. In W. Bos, E. Klieme, & O. Köller (Eds.), *Schulische Lerngelegenheiten und Kompetenzentwicklung* (pp. 127–146). Münster, Germany: Waxmann.
- Artelt, C., Neuenhaus, N., Lingel, K., & Schneider, W. (2012). Entwicklung und wechselseitige Effekte von metakognitiven und bereichsspezifischen Wissenskomponenten in der Sekundarstufe. *Psychologische Rundschau*, *63*, 18–25.
- Artelt, C., & Schneider, W. (in press). Cross-country generalizability of the role of metacognitive knowledge for students' strategy use and reading competence. *Teachers College Record*.
- Artelt, C., Weinert, S., & Carstensen, C. H. (2013). Assessing competencies across the lifespan within the German National Educational Panel Study (NEPS) – Editorial. *Journal for Educational Research Online*, *5*(2), 5–14.
- Azevedo, R. (2007). Understanding the complex nature of self-regulatory processes in learning with computer-based learning environments: An introduction. *Metacognition and Learning*, *2*, 57–65.
- Baker, L. (1989). Metacognition, comprehension monitoring, and the adult reader. *Educational Psychology Review*, *1*, 3–38.
- Baumert, J., Bos, W., & Lehmann, R. (2000). *TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. Opladen, Germany: Leske+Budrich.
- Baumert, J., Heyn, S., & Köller, O. (1992). *Das Kieler Lernstrategien-Inventar (KSI)*. Kiel, Germany: Universität Kiel.
- Berendes, K., Fey, D., Linberg, T., Wenz, S. E., Roßbach, H.-G., Schneider, T., & Weinert, S. (2011). Kindergarten and elementary school. 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. 203–216). Wiesbaden, Germany: VS.
- Blossfeld, H.-P., von Maurice, J. & Schneider, T. (2011). *Grundidee, Konzeption und Design des Nationalen Bildungspanels für Deutschland* (NEPS Working Paper No. 1). Bamberg, Germany: Otto-Friedrich-Universität, Nationales Bildungspanel.
- Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, *7*, 161–186.
- Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction*, *20*, 372–382. doi: 10.1016/j.learninstruc.2009.03.002
- Borkowski, J. G., Chan, K. S. L., & Muthukrishna, N. (2000). A process-oriented model of metacognition: Links between motivation and executive functioning. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition (Buros-Nebraska Symposium on Measurement & Testing)* (pp. 1–41). Lincoln, NE: Buros Institute of Mental Measurements.
- Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert, & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65–116). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., & Bandura, A. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology*, *100*, 525–534.

- Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review*, 20, 391–409.
- Dresel, M., & Haugwitz, M. (2008). A computer-based training approach to foster motivation and self-regulated learning. *Journal of Experimental Education*, 77, 3–20.
- Dubowy, M. (2010). *Metagedächtnisfähigkeiten und selbstbezogene Sprache bei Vorschulkindern – zwei verwandte Aspekte kindlicher Selbstregulation?* Berlin, Germany: Logos.
- Ebert, S. (2011). *Was Kinder über die mentale Welt wissen – Die Entwicklung von deklarativem Metagedächtnis aus der Sicht der “Theory of Mind”*. Bamberg, Germany: Otto-Friedrich-Universität Bamberg.
- Efkliides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, 46, 6–25.
- Ertmer, P., & Newby, T. (1996). The expert learner: Strategic, self-regulated and reflective. *Instructional Science*, 24, 1–24.
- Flavell, J. H. (1971). First discussant’s comments: What is memory development the development of? *Human Development*, 14, 272–278.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906–911.
- Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online*, 5(2), 50–79.
- Händel, M., Tupac-Yupanqui, A., & Lockl, K. (2012). *Metakognitives Wissen und der Einsatz von Lernstrategien bei Studierenden* (NEPS Working Paper No. 20). Bamberg, Germany: Otto-Friedrich-Universität, Nationales Bildungspanel.
- Hasselhorn, M. (2006). Metakognition. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (2nd ed., pp. 466–471). Weinheim, Germany: Beltz PVU.
- Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging*, 24, 462–475.
- Jacobs, J. E., & Paris, S. G. (1987). Children’s metacognition about reading: Issues in definition, measurement, and instruction. *Educational Psychologist*, 22, 255–278.
- 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, 1–60.
- Lockl, K. (2012). *Assessment of declarative metacognition: Starting Cohort 4 – Ninth Grade*. Bamberg, Germany: Otto-Friedrich-Universität, Nationales Bildungspanel.
- Lockl, K., Händel, M., Haberkorn, K., & Weinert, S. (2013). *Metacognitive knowledge in young children: Development of a new test procedure for first graders*. Manuscript submitted for publication.
- Lockl, K., & Schneider, W. (2002). Developmental trends in children’s feeling-of-knowing judgements. *International Journal of Behavioral Development*, 26, 327–333.
- Lockl, K., & Schneider, W. (2003). Metakognitive Überwachungs- und Selbstkontrollprozesse bei der Lernzeiteinteilung von Kindern. *Zeitschrift für Pädagogische Psychologie*, 17, 173–183.
- Maag Merki, K., Ramseier, E., & Karlen, Y. (in press). Reliability and validity analyses of a newly developed test to assess learning strategy knowledge. *Journal of Cognitive Education and Psychology*.
- Maki, R. H., Schields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, 97, 723–731.

- Marschner, J. (2011). *Adaptives Feedback zur Unterstützung des selbstregulierten Lernens durch Experimentieren*. Duisburg/Essen, Germany: Universität Duisburg-Essen.
- Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, 18, 159–163.
- Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, 94, 249–259.
- Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.
- Myers, M., & Paris, S. G. (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology*, 70, 680–690.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, 26, 125–141.
- 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.
- 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.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Michigan, WI: National Center for Research to Improve Postsecondary Teaching and Learning.
- Pintrich, P. R., Wolters, C., & Baxter, G. (2000). Assessing metacognition and self-regulated learning. In G. Schraw, & J. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Lincoln, NE: Buros Institute of Mental Measurements.
- Roßnagel, C. S., Bittner, J., & Staudinger, U. M. (2009). *Self-regulation across the lifespan: Its conceptualization and assessment in the National Educational Panel Study (NEPS): Expert Report*.
- Scherer, R., & Tiemann, R. (2012). Factors of problem-solving competency in a virtual chemistry environment: The role of metacognitive knowledge about strategies. *Computers & Education*, 59, 1199–1214.
- Schlagmüller, M., & Schneider, W. (2007). *WLST 7-12. Würzburger Lesestrategie-Wissenstest für die Klassen 7 bis 12*. Göttingen, Germany: Hogrefe.
- Schmitt, M. C. (1990). A questionnaire to measure children's awareness of strategic reading processes. *The Reading Teacher*, 43, 454–461.
- Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *International Mind, Brain, and Education Society*, 2, 114–121.
- Schneider, W., & Artelt, C. (2010). Metacognition and mathematics education. *The International Journal on Mathematics Education*, 42, 149–161.
- Schneider, W., & Lockl, K. (2006). Entwicklung metakognitiver Kompetenzen im Kindes- und Jugendalter. In W. Schneider, C. F. Graumann, N. Birbaumer, & B. Sodian (Eds.), *Enzyklopädie der Psychologie: Vol. 2. Entwicklungspsychologie* (pp. 721–767). Göttingen, Germany: Hogrefe.
- Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Handbook of metacognition in education* (pp. 415–429). New York, NY: Routledge.
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19, 460–475.
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7, 351–371.

- Schwinger, M., Laden, T. von der, & Spinath, B. (2007). Strategien zur Motivationsregulation und ihre Erfassung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *39*, 57–69.
- Spörer, N., & Brunstein, J. C. (2006). Erfassung selbstregulierten Lernens mit Selbstberichtsverfahren. *Zeitschrift für Pädagogische Psychologie*, *20*, 147–160.
- Statistisches Bundesamt. (2011). *Fachserie 11 Reihe 1. Bildung und Kultur. Allgemeinbildende Schulen*. Wiesbaden, Germany: Statistisches Bundesamt.
- Stumm, S., Thomas, E., & Dormann, C. (2010). Selbstregulationsstärke und Leistung. *Zeitschrift für Arbeits- und Organisationspsychologie*, *54*, 171–181.
- Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology*, *82*, 306–314.
- Thillmann, H. (2008). *Selbstreguliertes Lernen durch Experimentieren: Von der Erfassung zur Förderung*. Duisburg/Essen, Germany: Universität Duisburg-Essen.
- van Kraayenoord, K., & 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.
- van Velzen, J. H. (2012). Teaching metacognitive knowledge and developing expertise. *Teachers and Teaching*, *18*, 365–380.
- Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und Metakognition. Implikationen für Forschung und Praxis* (pp. 77–95). Münster, Germany: Waxmann.
- Veenman, M. V. J., & Elshout, J. J. (1999). Changes in the relation between cognitive and metacognitive skills during the acquisition of expertise. *European Journal of Psychology of Education*, *14*, 509–523.
- Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task. *Learning and Individual Differences*, *15*, 159–176.
- Veenman, M. V. J., van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*, 3–14.
- Vlachopoulos, S. P., & Efklides, A. (2012). Measurement of metacognitive knowledge of self, task, and strategies in mathematics. *European Journal of Psychological Assessment*, *28*, 227–239.
- 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.
- Weinstein, C. E., & Palmer, D. (1990). *LASSI-HS user's manual*. Clearwater, FL: H & H Publishing.
- Wirth, J., & Leutner, D. (2008). Self-regulated learning as a competence: Implications of theoretical models for assessment methods. *Zeitschrift für Psychologie*, *216*, 102–110.
- Wohlkinger, F., Ditton, H., Maurice, J. von, Haugwitz, M., & Blossfeld, H.-P. (2011). Motivational concepts and personality aspects across the life course. 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. 155–168). Wiesbaden, Germany: VS.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183. doi: 10.3102/0002831207312909

Appendix

Figure A1: Structure of the items. The item displayed is not part of the final item pool

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"/>