Heydrich, Jana; Weinert, Sabine; Nusser, Lena; Artelt, Cordula; Carstensen, Claus H.

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 5 (2013) 2, S. 217-240

urn:nbn:de:0111-opus-84315

Journal for Educational Research Online
Journal für Bildungsforschung Online
Volume 5 (2013), No. 2, 217–240

Jana Heydrich, Sabine Weinert, Lena Nusser,
Cordula Artelt & Claus H. Carstensen

Including students with special educational needs into large-scale assessments of competencies: Challenges and approaches within the German National Educational Panel Study (NEPS)

Abstract

The National Educational Panel Study (NEPS) as a newly set up large-scale assessment study in Germany has accepted the challenge of including students with special educational needs (SEN) into its conceptual design. Particularly, students with SEN in the area of learning (SEN-L) are oversampled within the NEPS. Their educational biographies and relevant context factors will be assessed longitudinally based on interviews and questionnaires given to their parents, teachers, and school principals. However, obtaining data (test data, questionnaires) from the target subjects themselves is by no means a simple, straightforward endeavor but requires careful research strategies.

In this article we will briefly discuss problems of specifying the target population of students with SEN-L and present a focused review of research literature relevant to the inclusion of students with SEN-L into large-scale assessments. Specifically, we will focus on challenges relating to the standardized, reli-

Jana Heydrich (corresponding author), National Educational Panel Study (NEPS), University of Bamberg, Wilhelmsplatz 3, 96045 Bamberg, Germany
e-mail: jana.heydrich@googlemail.com

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

Dipl.-Päd. Lena Nusser, National Educational Panel Study (NEPS), University of Bamberg, Wilhelmsplatz 3, 96045 Bamberg, Germany
e-mail: lena.nusser@uni-bamberg.de

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

Prof. Dr. Claus H. Carstensen, Chair of Psychology with a Focus on the Methods of Empirical Educational Research, University of Bamberg, Wilhelmsplatz 3, 96045 Bamberg, Germany
e-mail: claus.carstensen@uni-bamberg.de

able, and valid testing of competencies of students with SEN-L within large-scale assessments. Additionally, the article outlines the basic design of feasibility studies within the NEPS to test for the structural comparability of competence assessments in special schools with those in regular schools. These studies will further explore the necessity of test accommodations for students with SEN-L and the impact of those accommodations on the validity and comparability of test scores of students with and without SEN-L.

Keywords
Special educational needs; Measuring competencies; Accommodations; Feasibility study

Die Einbeziehung von Förderschülern in Large-Scale-Kompetenzerhebungen: Herausforderungen und Vorgehen im Rahmen des Nationalen Bildungspanels (NEPS)

Zusammenfassung
Das Nationale Bildungspanel (NEPS) als eine Large-Scale-Studie in deren Rahmen ein längsschnittliches Multi-Kohorten-Sequenz-Design zur Erforschung von Bildungsverläufen in Deutschland realisiert wird, bezieht im Kontext eines Oversamplings auch Schülerinnen und Schüler mit sonderpädagogischem Förderbedarf mit in ihr Erhebungsdesign ein. Der Fokus liegt dabei zunächst auf der größten Untergruppe von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf, nämlich Personen mit Förderbedarf im Bereich Lernen. Durch die Befragungen von Eltern, Lehrern und Schulleitern können deren Bildungswege unter Berücksichtigung relevanter Kontextbedingungen – erstmalig – anhand einer größeren Stichprobe nachgezeichnet und analysiert werden. Eine besondere Herausforderung stellt allerdings die Testung und Befragung der Probanden mit besonderem Förderbedarf im Bereich Lernen selbst dar.

Im vorliegenden Beitrag wird ausgehend von einer kurzen Diskussion der Stichprobendefinition ein fokussierter Überblick über Forschungsbefunde gegeben, welche für die Berücksichtigung von Probanden mit dem Förderschwerpunkt Lernen in Large-Scale-Studien besonders relevant sind. Auf dieser Basis werden die Herausforderungen erörtert, die der Einbezug von diesen Schülerinnen und Schülern vor allem mit Blick auf die standardisierte, reliable und valide Messung von bildungsbezogenen Kompetenzen darstellt. Es wird anschließend das Design von Machbarkeitsstudien innerhalb des NEPS skizziert. Diese Studien prüfen gezielt die strukturelle Vergleichbarkeit von Kompetenztestungen an Förderschulen mit jenen an Regelschulen sowie die Notwendigkeit und die Wirkungen von Adaptationen bei den Testungen mit Blick auf die Validität und Vergleichbarkeit von Testergebnissen bei Kindern mit und ohne sonderpädagogischen Förderbedarf im Bereich Lernen.

Schlagworte
Förderschüler; Kompetenzmessung; Test-Modifikation; Machbarkeitsstudie

1. Introduction

The population of students with special educational needs (SEN) comprises more than 485,400 students in Germany, which is around 6.4% of the entire student population (KMK, 2012, p. XI). Obviously, children with SEN do not comprise a homogenous but rather a heterogeneous group of students including, for example, students with learning disabilities, students with language disabilities, and students with physical handicaps/disabilities (cf. KMK, 1994, p. 6). In Germany, about 78% of these students (KMK 2012, p. XIII) do not attend regular schools, but instead attend special schools with specific schooling programs and trainings tailored to those students who appear to be unable to follow school lessons and subject matters in regular classes.

Yet, since the ratification of the UN-Convention in 2009, the lawmakers of the federal states of Germany have obliged to the effort of establishing an inclusive school system and to advocating more *integration*[1] in all federal educational institutions (BMAS, 2010, article 24). Thus, the integration rate is now growing every year: In 2001, 12.4% of all students with SEN attended regular schools; in 2010, we find 22.3% in regular school settings (KMK, 2012, p. XIII). Nevertheless, the percentage of students with SEN at special schools has remained almost consistent over the last years (4.9% of all students attend special schools, cf. KMK, 2012, p. XII). This suggests a growing rate of students with SEN in total which might be due to different reasons, for example changing criteria or a growing population.

The question of whether and to what extent special schools do in fact provide better learning opportunities and individual support with more positive consequences for the individual life course or, on the contrary, whether and to what extent they lead to stigmatization, exclusion, and extenuated chances on the labor market later on in life has been a matter of controversial debates among politicians, stakeholders as well as within the scientific community.

At the same time, regarded from a research perspective, we know comparatively little about the educational and vocational careers of this group of students. Even less is known about the development of competencies across the lifespan within the total group or subgroups of children with SEN. This is partially due to the fact that for a long time larger groups of students with SEN were hardly included into national or international large-scale assessments (cf. Hörmann, 2007; von Stechow, 2006). Although this has changed within the last years, an adequate database to close this research gap is still missing. Therefore, the National Educational Panel

1 In this article the term *integration* is used to express that students with and without SEN attend the same class at any type of school of the regular school system in Germany (Werning & Löser, 2010).

Study (NEPS) as a newly set up large-scale assessment study in Germany has accepted the challenge of including students with SEN into its conceptual design.

Since students with SEN comprise a heterogeneous group posing different challenges when being included in a large-scale assessment study, as a first approach NEPS focuses on the numerically largest group of students with SEN, that is, students with special needs in the area of learning (SEN-L). In Germany, this subgroup comprises around 202,200 students, which is 41.6% of all students with SEN (KMK, 2012, p. XI). In 2010 most of them attended special schools: While about 13% of these students can be found at regular primary schools (calculation based on the KMK, 2012, p. 65ff), this percentage declines when it comes to secondary education. At this educational stage just about 8.7% of all students with SEN-L attended regular school settings (calculation based on KMK, 2012, p. 77ff). Within the first funding period of the NEPS, the oversampled group of students with SEN-L was sampled at special schools only.

Choosing students with SEN-L for detailed feasibility studies on the validity of competence assessments as well as for gathering in depth background information reported by their parents, teachers, or school principals is of special importance when considering both, the frequency of students with SEN-L in the population and the fact that in 2010 about 66% of all students who left school without a regular qualification certificate after the period of compulsory school attendance, have attended a special school for students with SEN-L (KMK, 2012, p. 45f).

As indicated, data obtained from the target persons' parents, teachers, and school principals will provide information on educational careers and relevant context variables including home background, formal and informal learning opportunities, educational decisions, and educational returns. This data will be released via various access modes to the national and international scientific community (see Blossfeld, von Maurice, & Schneider, 2011). On this basis the following important research questions can be addressed:

- Which educational and vocational careers are to be expected in groups of students with and without SEN-L who achieve no qualification certificate?
- How important are the obtained certificates for individual careers, chances on the labor market, and personal development in later life?
- Which variables influence educational decisions in both groups – students with SEN-L in comparison to those without SEN-L?
- Which aspects (learning environments, social, and economic circumstances) promote or constrict the attainment of educational goals?

However, obtaining valid data from the target subjects with SEN-L themselves within large-scale assessment studies that allow for comparisons with regular school students is by no means a simple, straightforward endeavor. Instead, it requires systematic research on the effects of varying kinds of accommodations of the testing conditions to the needs of students with SEN-L on the validity and comparability of the derived competence indicators within and across school types. In

this paper we discuss these challenges and derive a set of feasibility studies to address the following questions empirically:

a) Can valid comparisons between students with SEN-L and regular school students be made on the basis of NEPS competence tests which are targeted at students in regular schools?
b) Which accommodations can and should be implemented at special schools to obtain valid competence measures from students with SEN-L within the NEPS?
c) How do these accommodations influence performances and competence scores of students attending regular school settings, in particular lower performing students attending *Hauptschule*.[2]

In the following sections we will briefly discuss the difficulty of defining SEN-L and present a focused review on some large-scale research experience associated with this target group. Furthermore, the interaction between standardization and modification of testing, some formal affordances in large-scale studies, and the consequences for obtaining valid and comparable test scores will be discussed. Before going into detail on the feasibility studies relating to students with SEN-L within the NEPS, we will provide some basic information about the study as a whole. A summary of the potentials and limitations of our project will conclude our argument.

2. Students with SEN-L: Definition of the target population

To start with the main conclusion, there is no precise, consistent, and generally accepted definition of SEN-L in the national and international research literature (cf. Bleidick, 1968b; Eberwein, 1997; Hammill, 1990). Within different countries as well as within different academic disciplines (e.g., Germany, USA; psychology, educational science) varying descriptions of the phenomenon of "special educational needs in the area of learning" (SEN-L) have evolved and have been covered by a variety of terms such as *Lernbehinderung* (cf. Kanter, 1997), *Lernschwierigkeiten* (cf. Zielinski, 1995), *Lernstörung* (cf. Lauth, Brunstein, & Grünke, 2004) in German speaking countries and *learning disabilities*, *learning difficulties* (LD) in the United States (cf. U.S. Department of Education, 2004), or *specific learning difficulty* (SpLD) in the United Kingdom.

Most notably the German term *Lernbehinderung* does not depict the same concept as the English terms *learning difficulties* or *learning disabilities* (Schröder, 2000). Contrary to the conceptualization of the German term *Lernbehinderung,* the Anglo-American concept of *learning disabilities* is much broader and sub-

2 The *Hauptschule* is a type of school at lower secondary level providing a basic general education. It is compulsory for all students not attending a different type of secondary school, usually comprising grades five to nine (KMK, 2009).

sumes not only *mild mentally/intellectually retarded* or *educable mentally retarded* students (i.e., students or adults comprising an under-average IQ of 50 to 75) but also children with *learning disabilities* (i.e., students performing within normal range on standardized IQ-tests but not at school; so-called "underachievers") (cf. Schröder, 2000).

In the same vein, the German term *Lernschwierigkeiten* covers a broad range of phenomena. Zielinski (1995, p. 13), citing Weinert and Zielinski (1977), provides the following widely accepted functional definition of the phenomena denominated by this term:

> Talking about Lernschwierigkeiten in general means that a student's school achievement is below a tolerable discrepancy to obligatory institutional, social, and individual benchmarks (like standards, requirements, expectations), or that the reaching (or missing) of the respective standard is associated with burdens, which lead to inacceptable side effects in behavior, experience, or in the development of personality of the learner. (authors' translation; Zielinski, 1995, p. 13)

Especially within the German school system the requirements of the various types of schools may differ tremendously. Nevertheless, *Lernschwierigkeiten* may be observed in every type of school. In addition, depending on the respective benchmark (individual, institutional, social norm) for evaluating students' performances and achievements, quite different groups of children and/or adults are focused on. Social or institutional benchmarks specify discrepancies relative to either the total age group or to the classmates, the students of the same grade or the same school type respectively. Discrepancies relating to an individual norm refer to heterogeneous performance profiles of the same individual, or changing performances of an individual over time (cf. Zielinski, 1995).

Hence, based on their definition, the authors suggest various classification systems to characterize different types of *Lernschwierigkeiten* along with the abovementioned kinds of benchmarks, for example, the classification of subgroups according to the affected learning content, subject or domain, or, according to the degree of deviation from one or the other norm.

If *Lernschwierigkeiten* are prominent in many school-relevant subjects and persist over a long time, Zielinski (1995) and Lauth et al. (2004) talk of *Lernbehinderung*, that is, the mode of *Lernschwierigkeiten* which may require an allocation to a special school. In a slightly different vein, Schmetz (1999), referring to institutional norms, talks of *Lernbehinderung* as an expression denoting the failure and/or missing fit between a child's individual learning potential and the normative expectations of the regular school system.

In Germany labeling students with the term *Lernbehinderung* is still a common practice. It has evolved over time for the purpose of better supporting these students by allocating them to special schools (Eberwein, 1997). Nevertheless, many authors have come to the conclusion that nowadays *Lernbehinderung* should only be used as a working term for the existing system of selection in Germany

(cf. Schröder, 2000; Kanter, 2001). Even though decisions about a child's special needs in the area of learning (classifying them as *lernbehindert* or comprising a *Lernbehinderung*) are made by trained experts using diverse psychological and diagnostic tests, the classification criteria are not clear and rather questionable considering that there is no common definition of the phenomenon. In particular, as outlined by ecological and contextual models of development, individual characteristics and performances emerge as a consequence of the interaction of individual prerequisites and environmental stimulation and support in the course of development. As a consequence, in special education as well as in psychology *Lernbehinderung* or SEN-L is no longer monocausally defined as a lack of intelligence (Bleidick, 1968a); instead research has turned to a more systemic, multifactorial or contextual approach since the late 1970s (cf. Eberwein, 1997; Antor & Bleidick, 2001).

Thus, when considering students with SEN and with SEN-L we are to expect a highly heterogeneous population. Over and above distinguishing between different types of SEN, students diagnosed as SEN-L may comprise widely heterogeneous competence profiles (between subjects and within each subject). These profiles depend on the internal, external, and moderating conditions that have influenced their acquisition of abilities, skills, and knowledge structures as well as their development of motivational, emotional, and behavioral dispositions that influence learning and performance at school. Most notably the heterogeneous competence profiles and characteristics of these students may challenge test validity and thus may explain why researchers have not put too much effort into including reasonable sample sizes of this target population into large-scale assessments until now.

3. Students with SEN in existing large-scale assessments – A focused review

Overall, most large-scale assessment studies include students with SEN in their sample, however not in a substantial number. Hörmann (2007) expresses this very tellingly in the title of her graduation thesis, calling it "The invisibles in PISA, TIMSS & Co" (authors' translation). These large-scale assessment studies do not have the agenda to explore students with SEN in particular. The small number of students with SEN that have been included in these samples are mainly involved for reasons of representativeness.[3] They were usually surveyed with shortened measurement instruments, fewer tasks, and/or less testing time, which narrows the possibility to compare results of students with SEN-L to results of other groups of students. That is, because students with SEN-L who attend regular classes do have to cope with the longer and more difficult – regular – test version which is possi-

3 Included sample sizes of students with SEN in PISA: 108 students in 2003 (Prenzel, Drechsel, Carstensen, & Ramm, 2004); 160 students in 2006 (Prenzel, Carstensen, Frey, Drechsel, & Rönnebeck, 2007); 179 students in 2009 (Jude & Klieme, 2010).

bly leading to underestimation of their competence as compared to students with SEN-L in special schools processing the reduced version. Furthermore, since PISA data are cross-sectional and represent a specific age group, it is not possible to follow a student's educational career.

Within the German *Panel Study at the Research School "Education and Capabilities" in North Rhine-Westphalia* (PARS), students with SEN-L were sampled for explicit analysis of special needs education (PARS-F) (Institut für Schulentwicklungsforschung & Research School Education and Capabilities, n.d.). Due to various reasons the researchers decided to use specially designed measurement instruments for students with SEN-L. Hence, although they were able to generate a large-scale database, again, there will be no possibility to compare the assessments of students with SEN-L to students attending regular schools. Another restriction of this study is the limited geographical area of data collection which focuses on only a single federal state in Germany: North Rhine-Westphalia.

Although a comprehensive, representative, large-scale study including students with SEN-L is still missing; some interesting research projects have already been conducted aiming to include students with SEN into large-scale assessments and to compare test scores of students with and without SEN.

Such an effort was undertaken by Wocken (2000) within the so called LAU-F study[4] which was conducted in Hamburg. In this study Wocken adopted the measurement instruments administered in the LAU-study in Grade 5 for Grade 7 students with SEN-L attending special schools, thereby implementing the so called *out-of-level testing* method (cf. Koretz & Barton, 2003, p. 6). One key result was – as expected – the substantial deficit in cognitive performance (orthography and, although less profound, reasoning): 7th-grade students attending special schools still lagged behind nearly one standard deviation within the orthography test compared to the performances of 5th graders attending *orientation stage*[5] at lower secondary schools. Furthermore, Wocken compared the test performance of students with SEN-L at Grade 7 attending special schools to Grade 5 students attending regular schools who had similar cognitive preconditions as indicated by the CFT 20 raw scores (Culture Fair Intelligence Test; Weiß, 1978) (ex-post, quasi-experimental design) and still found significant deficits in orthography for the students with SEN-L. Unfortunately, Wocken did not report results on the validity and comparability of the implemented *out-of-level* testing.

4 LAU-F: "Untersuchungen zur Lernausgangslage an Förderschulen [Investigations of the Status Quo in Learning at Special Schools]". LAU means "Untersuchungen zur Lernausgangslage" and aimed to measure performances of all Grade 5 students, their learning progress and attitudes and follow them through their school careers (Lehmann & Peek, 1997).

5 The *Orientation stage* – either Grades 5 and 6 at the individual lower secondary school types or, in some Länder, is an independent school stage not attached to any school type. The orientation stage helps to decide on a pupil's future school career (KMK, 2009).

In 2005, Wocken continued his research at special schools in Hamburg drawing on the study called KESS 7[6] (Bos, Bonsen, & Gröhlich, 2009). The KESS-7-F study (Wocken & Gröhlich, 2009) provided the opportunity to compare 7th graders of special schools with 4th-grade students attending regular school settings by implementing items from KESS-4 competence tests into the KESS-7-F competence testing (*anchor-item design*, cf. Wocken & Gröhlich, 2009, p. 98). Results, applying Item Response Theory this time, demonstrated a performance discrepancy of more than two years between both groups. Specifically, 4th-grade primary school students outperformed students with SEN-L in 7th grade in reading competence, the difference being about 1/3 of a standard deviation; in mathematics the 7th-grade students with SEN-L scored 2/3 of a standard deviation below their younger peers without SEN. The findings of the implemented subtest "figural reasoning" of the cognitive abilities test (*Kognitiver Fähigkeitstest* (KFT); Heller & Perleth, 2000) concurred with previous results: Again, students with SEN-L appeared to be cognitively disadvantaged when taking into account their raw scores in comparison to those of their peers at Grade 7 attending regular schools (cf. Wocken & Gröhlich, 2009).

Further results were obtained by the LABEL 8-10[7] and the BELLA study[8], conducted in Berlin in 2002 and 2006 (cf. Lehmann, Nikolova, & Peek, 2004; Lehmann & Hoffmann, 2009). Drawing on the LABEL results the BELLA study offered the possibility to contrast students with SEN-L in different tracks of schooling. Contrary to the results of Wocken (2000), the BELLA study did not find any significant differences in students' performances depending on their school tracking, that is, between matched students with SEN-L at regular or at special schools. Another interesting finding of BELLA is that competencies of students with SEN-L did not show normal distributions of test scores in different domains but often very broad, bimodal or multimodal distributions, exposing the possible existence of subgroups of students with SEN-L.

To summarize, various studies focusing on students with SEN-L have been conducted in Germany in recent years using different measurement instruments, different kinds of test accommodations, methods, and different modes of administration when exploring various competence domains in students with SEN-L. For all that, none of these studies can be considered as being representative for Germany or as comprising a coherent longitudinal design, but rather as a continuation of locally restricted studies. In particular, most of these studies did not offer the possi-

6 KESS: "Kompetenzen und Einstellungen von Schülerinnen und Schülern [Competencies and Attitudes of Students]"; KESS 7 was conducted in Hamburg including 7th-grade students of all secondary schools. KESS-7-F surveyed seventh-grade students attending special needs schools.

7 LABEL 8-10: "*Lernausgangslage arbeitsrelevanter Basiskompetenzen im Förderschwerpunkt Lernen in Klassen 8-10* [Learning Foundations for Work-Relevant Basic Competencies of Students with SEN-L in Grades 8 to 10]".

8 BELLA: "*Berliner Erhebung arbeitsrelevanter Basiskompetenzen von Schülerinnen und Schülern mit Förderschwerpunkt Lernen* [Berlin Survey of Work-Relevant Basic Competencies of Students with SEN-L]".

bility to compare the development of students with SEN-L at special schools with developmental progress seen in students attending regular schools. Furthermore, there were hardly any hints to validity and comparable item functioning for the implemented methods, although it was assumed that the same constructs were being assessed.

4. Including students with SEN-L into large-scale assessments of competencies in general: Threats to validity and comparability of test scores

Precursory research on students with SEN-L shows that measuring competencies in this heterogeneous group of children validly and allowing for comparisons between groups of children with and without SEN-L is an ambitious task. With respect to large-scale assessments like the NEPS survey, without empirical evidence on its effects, it is neither sufficient to simply administer the same test instruments to different populations of students nor to accommodate tests and test administration to students attending special schools. Both strategies may lead to insufficient test fairness and could subsequently yield invalid data. To reasonably compare test performances or competence parameters across groups, methodological considerations and quality criteria of psychological testing must be taken into account and the validity and comparability of measurements obtained has to be explored explicitly.

4.1 Issues of standardization, accommodation, and validity of assessments

To allow for comparisons between groups a valid database of assessments of the same concepts is required. From the perspective of measurement, a precondition for valid data is the standardization of testing, for example, standardized instructions, standardized measurement instruments, and scoring criteria. In the case of students with SEN, however, using comparable standardized testing procedures might itself distort measurement and interpretations by negatively biasing test scores because of – construct-irrelevant – disability-related impediments to performance (cf. Koretz & Barton, 2003, p. 6f). To correct for the implemented disadvantages of students with SEN in school assessments, a corollary is to provide modifications of testing in order to increase test fairness and the validity of information about those participants. In the current edition of *Americas Standards for Educational and Psychological Testing* the term *accommodations* is suggested to be used

> [...] as the general term for any action taken in response to a determination that an individual's disability requires a departure from established testing

> protocol. Depending on circumstances, such accommodation may include modification of test administration processes or modification of test content. (AERA, 1999, p. 110)

4.1.1 Test administration

The National Center of Educational Outcomes (NCEO) in the U.S. specifies four categories of accommodations concerning modifications of test administration: (a) accommodation of presentation (e.g., read directions to students, provide audiotaped administration of sections, read questions aloud to student), (b) accommodation of time/scheduling (e.g., extend the time allotted to complete the test, allow frequent breaks during the testing), (c) accommodation in student's response (e.g., provide wider lines, provide word processor, provide copy assistance between drafts), and (d) accommodation of test setting (e.g., provide special lighting, special furniture, allow testing in a small group) (Elliott, Thurlow, & Ysseldyke, 1996, p. 4f). These accommodations provide appropriate methods for assessing students when their disability mostly relates to sensory or physical handicaps. McDonnell, McLaughlin, & Morrison (1997) introduced the metaphor, that:

> Accommodations are intended to function as a corrective lens that will deflect the distorted array of observed scores back to where they ought to be – that is, back to where they provide a more valid image of the performance of individuals with disabilities. (McDonnell et al., 1997, p. 176)

The issue becomes more complicated when students' ability profiles have a more direct impact on test performance, that is, if specific disabilities that are not in the focus of the assessment impair test performance and are therefore relevant to test validity. For example, "[...] performance assessments in the area of mathematics are likely to involve reading and writing [...] this [...] increases the probability that reading or writing disabilities, which are among the most common, will interfere with the assessment of mathematics" (McDonnell et al., 1997, p. 171).

The most common modifications in test administration provided in school assessments in the U.S. are *extended time* and *reading items or directions aloud* (cf. Koretz & Barton, 2003, p. 15). How these accommodations affect survey findings is documented by the NCEO and the American Educational Research Association (AERA). Overall, only a few consistent conclusions can be drawn across studies. In the following some of these will be discussed with a special focus on large-scale assessments like the NEPS.

4.1.1.1 Testing time within large-scale studies. Competence tests within large-scale studies like the NEPS are often composed as *speed-and-power-tests*. While the individual working time allocation is allowed within a specific test, the total test time is fixed, that is, achievement is conceptualized as performance within time limits. In the NEPS the test time for each of the competence domains is fixed to

around thirty minutes. Especially when it comes to the implementation of different subtests administered to groups of students or the combination of tests and questionnaires, fixed test times are indispensable for practical reasons. In addition, time limits are construct-relevant. Obviously, as Kubinger (2009) notes, individual scores of speed-power tests may be traced back to either low ability or low working speed. The issue of test time is even more complicated in heterogeneous groups of children who may or may not comprise a reduced information processing speed. In fact, presenting students with more time does not have a single, but possibly many different effects depending on (a) the individual speed of information processing relevant to the task and competence domain under study (e.g., access to word representations in long-term memory), (b) the speed of conducting various other affordances of the respective task, and (c) the ability to sustain attention over longer periods of time.

Hence, it is not astonishing that there is no general valid finding for the impact of extended working time on test scores in students with SEN (Cormier, Altman, Shyyan, & Thurlow, 2010). However, across many studies extended time tended to improve the performance of all students (students with and without SEN), while students with *learning disabilities* tended to exhibit relatively greater score gains (cf. Sireci, Scarpati, & Li, 2005).

4.1.1.2 Reading items or directions aloud. Within large-scale assessments tests are predominantly administered to groups of students. Kubinger (2009) citing Amelang and Schmidt-Atzert (2006) highlights the expected disadvantages for bad readers in group testing due to the reading burden of written tasks. Reading the tasks and items to the children may reduce this disadvantage. Yet, not much is known about how this modification changes the processing of the task and the individual allocation of study time and how these changes in turn affect test performance in the individual child. With respect to students with SEN existing empirical evidence shows, for example, that oral accommodations on math tests were associated with increased test performance for some but not all students with disabilities (cf. Sireci et al., 2005).

4.1.2 Demands of the measurement instrument itself

With respect to accommodations regarding specific aspects and demands of the measurement instrument itself, three types of accommodations have to be distinguished: (a) design-related, (b) content-related, and (c) construct-related accommodations. The prevalent procedure in German studies, for example, PISA and KESS-7-F, including students with SEN-L was to skip some of the subtests; mostly in order to keep the burden for the special targets low, which is a design-related accommodation (e.g., cf. Bos et al., 2009). This accommodation is expected to have low or even no impact on the content of the administered subtests themselves. An already mentioned – and in matters of students with SEN-L apparently common –

practice is the use of *out-of-level* testing and *anchor-item* design. These approaches do not have any influence on the assessed construct per se but have an effect on the content of measurement instruments, for instance, when using construct-related tasks developed for younger students. A third accommodation regarding special aspects and demands of the implemented test instruments themselves is to leave out special types of requirements, for example, all tasks drawing on curriculum-based knowledge background which has not been taught at special schools due to their more restricted curriculum. If these requirements tap the intended construct to be measured, the accommodation becomes "construct-relevant" (cf. Koretz & Barton, 2003, p. 9f).

4.1.2.1 Item-format. Within large-scale studies, maintaining criteria of objectivity – an indispensable precondition for obtaining interpretable and valid data – affords administration modes and test conditions that are rather easy to implement without much variance between administrators. It is important to take great care that standardization and the burden of testing do not result in selective panel mortality in future survey waves. Large-scale studies have to adjust to testing large samples and, hence, have to consider economic aspects pertaining to resources as well as testing time and scoring efforts. On these grounds, tests are mainly presented to groups of students and in a closed response format (multiple choice [MC]) allowing for an easy and objective scoring of performances is preferred to open response formats.

With respect to item format, a literature review shows that students with SEN do not seem to have those specific skills available that are described in the literature as *test-ability* or *test-wiseness*. Meyerhöfer (2007) specifies components of *test-ability* or *test-wiseness* as follows: timing strategies, strategies to avoid mistakes, strategies of guessing answers, strategies of using hints for finding the right answer. He concludes that *test-ability* is always involved when it comes to a MC-format and guessing right answers. Students with SEN-L do not seem to have these metacognitive skills as readily available as their peers (cf. Lauth, Brunstein, & Grünke, 2004, p. 15f). This fact forces them into a disadvantage, which makes it hard to guarantee full test fairness. In fact, Johns and Vanleirsburg (1992) showed that students with SEN benefit from training in test-taking by scoring significantly higher after training. Nevertheless, designing tasks in MC-format is still of advantage when it comes to large sample sizes. This is also because it provides less demand in productive writing, which is often a limited ability of students with SEN-L as well.

To sum up, important reasons for using accommodations are to increase the validity of measurement in students with special educational needs, that is, to assure that test scores refer to those constructs that are intended to be assessed (cf. Koretz & Barton, 2003). In addition, by implementing accommodations the participation rates in national and state assessments can be increased, as shown in the U.S. (cf. NCLD, 2005, p. 2).

4.2 Comparability of test data

Modifications concerning test content always have direct impact on comparability between groups tested with or without the respective accommodation. In particular, when implementing accommodations within special schools for children with SEN-L, comparability of test scores obtained for these children becomes arguable unless we do know how students of regular schools tested without accommodations would perform when offered the same modifications. In fact, accommodations offered to just one group of students may distort group comparisons by potentially favoring students with SEN in special schools compared to students attending regular schools who are tested without accommodations, just as testing SEN-L students without accommodations may lead to an invalid underestimation of their competencies.

5. Including students with SEN-L into the NEPS: Potential and feasibility studies

In the following section, we derive a set of feasibility studies that were set up to explicitly test for the possibility of including students with SEN-L into the regular test program of the NEPS.

Before describing our approach concerning the feasibility studies in more detail, we will present a brief overview on some relevant features of the NEPS design and the assessments of competencies within NEPS.

5.1 Sampling, design, and measurement of competencies within the NEPS: A brief overview

5.1.1 Sampling and design

In the NEPS a longitudinal multicohort sequence design is implemented. Several cohorts of targets, namely infants, 4-year-olds attending kindergarten, 5th-grade students (10 to 11-year-olds), 9th-grade students (14 to 15-year-olds), 1st-year higher education students, and 23- to 64-year-old adults will be surveyed over the lifespan (cf. Blossfeld et al., 2011). In addition to sampling students in regular schools, about 600 students with SEN-L were sampled in Grade 5 and about 1,100 students in Grade 9 at special schools in 2010.

In general, the NEPS will allow for elaborate descriptions and analyses of the targets' long-term development and educational careers from five theoretically interconnected perspectives, namely from the perspective of (1) competence development across the lifespan, (2) education processes in learning environments, (3) social inequality and educational decisions over the life course, (4) education acqui-

sition of people with migration background, and (5) returns of education in the life course. Thus, within the NEPS life courses, competence development, and contextual influences are to be explored in their developmental interrelations drawing on information given by the target persons themselves and by context persons, such as parents, teachers, and school principals.

With respect to students with SEN-L, all context persons respond to the NEPS surveys on a regular basis. Hence, data related to students with SEN-L can be analysed by national and international researchers, via scientific use files for instance (for further information about the NEPS, i.e., concrete sample sizes and comprehensive research questions see Blossfeld et al., 2011). However, data obtained from the target persons with SEN-L themselves are subject to the feasibility studies outlined in the next section.

Note, that the NEPS sample recruited in regular schools may include some students with SEN by chance. However, this group of students will be numerically too small as to allow for systematic and meaningful comparisons between students with SEN-L attending special schools versus students with SEN-L attending regular schools.[9] Comparisons between students with SEN-L and other students in regular classes depend – as already argued – on the validity and comparability of the assessments.

5.1.2 Measuring of competencies in the NEPS

In educational contexts as well as in the NEPS, competencies are defined as functional, context-bound, domain-, and demand-specific (cognitive) achievement dispositions that are subject to educational influence and interventions (Weinert, 2001). In the NEPS, German-language competencies (reading competence and oral language comprehension), mathematical competencies, and natural science competencies are to be assessed consistently and coherently across the lifespan (Weinert, Artelt, Prenzel, Senkbeil, Ehmke, & Carstensen, 2011). The frameworks for assessing these domain-specific competencies are related to the conception of (functional) literacy as a predictor for successful participation in society (OECD, 2006). Thus, tests rely on everyday problems that are more or less distant from school curricula. Within the NEPS, fixed combinations of domain-specific competence tests are assessed every other year (in the kindergarten and school cohorts) in alternating sequences. The intervals between assessments will increase in older cohorts. In addition, a nonverbal "culture-fair" indicator of domain-general cognitive functioning is assessed (perceptual speed, figural reasoning) as well as indicators of metacompetencies (declarative and procedural metacognition, information and computer technology literacy) and social competencies, the latter being mea-

9 Therefore, an additionally implemented project within the NEPS is currently working on developing strategies for identifying students with diagnosed SEN in regular schools in order to allow for a systematic sampling of these students (considering an oversampling of this group) (Gresch, Leuze, & Solga, 2011).

sured by questionnaires. Finally, a selection of stage-specific skills and outcome measures (e.g., phonological awareness in kindergarten, orthography at secondary schools, and occupation-related attainments of vocational training) as well as foreign and first language competencies in persons with migration background are to be assessed within the NEPS at certain times. For more detailed information see Weinert et al. (2011) and Frahm et al. (2011), as well as the other articles in this issue.

5.2 Feasibility studies on the validity and comparability of competence measures obtained from students with SEN-L within the NEPS

In the NEPS various feasibility studies are conducted in order to investigate whether and how valid competence measures can be obtained from students with SEN-L. These quantitative feasibility studies were preceded by a series of qualitative prestudies aiming to get a preliminary impression of how students with SEN-L might in principle deal with the NEPS competence tests. Basing on the results of these qualitative studies as well as on the results of larger pilot studies at regular schools the series of feasibility studies for the assessment of competencies of students with SEN-L has been planned.

5.2.1 Qualitative prestudies

The qualitative prestudies were conducted to get some preliminary insights into the following questions: Do students with SEN-L understand the NEPS' test instructions and the vocabulary used in the survey instruments? How long does it take them to read the texts and items? Do they manage the different response formats of the NEPS reading test in principle? Which problems may occur during a group test session compared to individual test situations when surveying students with SEN-L?

These prestudies included students with SEN-L in Grade 5, 6, and 9 at special schools. More than 80 individual interviews were conducted and about 60 students with SEN-L participated in group sessions. A special focus was on reading time and understandability of texts and items from the NEPS reading competence tests (and on items from the students' questionnaires).

Overall, an important result from the qualitative prestudies was that the tests do not seem to imply a standard that – in principle – cannot be met by students with SEN-L and that no avoidance behavior of students with SEN-L was observed in these qualitative prestudies – neither in individual interviews nor in group test situations.

Over and above this important general conclusion, quite a few specific results and insights emerged. In particular, drawing on the interviews, some texts and

items could be identified as being specifically problematic in wording, grammar, and topic for this group of students (or at least for subgroups of them). In addition, specific response formats and comprehension requirements within the reading test were identified as particularly challenging for students with SEN-L as indicated by lots of requests for clarification or inadequate reactions (for further information about the NEPS reading competence test see Gehrer, Zimmermann, Artelt, & Weinert, 2013, this issue). Further, as expected, overall reading speed and fluency appeared to be rather low. The observations and interviews helped to improve the instructions to be given by the test administrators and to specify the accommodations to be tested in the experimental design of the feasibility studies presented in the next paragraph.

5.2.2 Research questions and design of the feasibility studies

In general, the quantitative feasibility studies address the research questions already mentioned, focusing on (1) the validity of competence-score comparisons between students with and without SEN-L when being assessed with the standard NEPS instruments, (2) which accommodations can and should be implemented at special schools to obtain valid competence measures from students with SEN-L within the NEPS, and (3) what are the effects of these accommodations on performances and competence scores of students attending regular school settings, particularly those students attending *Hauptschule*.

To address these questions, the two main-samples of students with SEN-L (Grade 5 and 9) will be assessed in a longitudinal experimental design. Specifically, within these samples booklets with different experimental variations of test instruments and testing conditions have been implemented (and are to be implemented in the next assessment waves). Thus, the effects of various accommodations that are potentially relevant to validity and comparability of competence scores (e.g., the effects of skipping difficult texts, of skipping difficult items, of test length, and of *out-of-level* testing, etc.) are explicitly tested within the two main-samples of students with SEN-L attending special schools as well as within an additional sample of about 600 students attending *Hauptschule* constituting the control group. The assignment of the different test booklets occurred randomly at individual level.

The basic design of the feasibility studies will be illustrated by the test design of Grade 5 and by the accommodations concerning the administered reading competence test (see Table 1).

Table 1: Design of the feasibility studies relating to the assessments of reading competence for students with SEN-L attending special needs schools in Grade 5. Booklets three to six were administered in a control group at *Hauptschulen* in Grade 5 as well

Booklet 1	Booklet 2	Booklet 3	Booklet 4	Booklet 5	Booklet 6
Reading speed test					
Reading competence test, regular school	Mathematical competence test	Reading competence test, *out-of-level and anchor item design*	Mathematical competence test	Reading competence test, *out-of-level and anchor item design,* rotated	Mathematical competence test
Mathematical competence test	Reading competence test, regular school	Mathematical competence test	Reading competence test, *out-of-level and anchor item design*	Mathematical competence test	Reading competence test, reduced requirement, and *anchor item design*
Break					
Student's questionnaire					

Note. Test booklets are randomly assigned at individual level.

Regarding the first research question, one subgroup of students with SEN-L was presented with the same test instruments as students attending regular schools (booklets one and two). From the empirical results of this subgroup it can be seen, how many items students with SEN-L work on in the working time given, whether their distributions indicate floor effects, and whether the standard errors of the competence scores will have acceptable values. Furthermore, it can be investigated whether the item difficulty parameters of students with SEN-L are consistent with those of the main study assessment of students attending regular schools; that is, whether measurement invariance can be assumed between students with and without SEN-L.

As a further issue, variations in the order of presentation of test domains or test items respectively have been implemented into the booklet design. Thus, we can get evidence on the time span students with SEN-L are able to sustain their attention, that is, whether their attention declines substantially after working on a 30 minute block of test items.

To address the second research question, another subgroup of students with SEN-L received accommodated versions of test instruments (booklets three to six). Specifically, to investigate the potential of *out-of-level* testing, respective conditions were implemented. Within all booklets, common items (*anchor-item* design*)* have been implemented allowing for direct comparisons between all experimental and quasi-experimental conditions. Except the two anchor texts and their respective test items these test booklets contain texts and items developed for Grade 3

students attending regular schools. Although content and item difficulty have been adapted in these booklets, the conceptualization of test requirements is equivalent to the regular Grade 5 reading test (see Gehrer et al., 2013, this issue). These conditions will allow testing for the effects of content-related accommodations; that is, whether this type of accommodation could lead to more precise and unbiased estimations of students' abilities. Note that experiences from other studies using the *out-of-level* testing did not provide clear pro or contra arguments for using this method. Nevertheless, this type of accommodation might show that the abilities of students with SEN-L could be aligned with the abilities of younger children in regular schools when being assessed with instruments allowing for a coherent assessment across ages. Apart from the theoretical perspective this would indicate a pragmatically important perspective on constructing test instruments for students with SEN-L in future assessments of the NEPS.

To investigate another partly content related accommodation, empirically proven difficult items and one of the five texts presented in regular testing have been deleted in a further condition (booklet six). The omitted text type requires weighting the author's arguments, which is a potentially high cognitive demand for students with SEN-L. Hence, this booklet does not comply entirely with the construct (construct-relevant accommodation) of reading competence within the NEPS. However, when reducing test requirements in order to accommodate for low reading skills we have to investigate the validity of the test instrument in this condition.

To address the third question, a special data collection was conducted. The booklets three to six have been administered in identical manner in students attending *Hauptschulen*. By comparing the results of students at regular schools with the same tests and accommodations we can analyze the interaction effects of the implemented accommodations between these children and students with SEN-L, and therefore examine the differential boost hypothesis (cf. Bolt & Ysseldyke, 2007, p. 125f).

Since students with SEN-L comprise a heterogeneous group and since meaningful comparisons to children without SEN depend on suitable matching procedures, some additionally selected measurement instruments have been implemented to gain more diagnostic information about the target population and the students' individual ability profiles. Standardized diagnostic tests to assess verbal and nonverbal reasoning, concentration, and attention were implemented (subtests of the "*Kognitiver Fähigkeitstest*" [KFT][10] and the "*Frankfurter Aufmerksamkeitsinventar"* [FAIR][11]). In addition, the NEPS indicator on domain-general cognitive functioning (nonverbal, culture-fair) and a NEPS test assessing reading speed have been administered. These measures should allow for describ-

10 KFT 4-12+R (Heller & Perleth, 2000): "Kognitiver Fähigkeitstest für Klassen 4-12, Revision [Cognitive Ability Test for 4th–12th grade, revised]"; used subtests: vocabulary, verbal reasoning, figural matrices and nonverbal reasoning.

11 FAIR (Moosbrugger & Oehlschlägel, 1996): "Frankfurter Aufmerksamkeitsinventar [Frankfurt Attention Inventory]"; measuring attention quality, global attention performance and continuity in attention.

ing the sample of students with SEN-L as well as the control group sample of students attending *Hauptschule*. Further, these diagnostic tests have been implemented to statistically control for heterogeneity in the analyses and to match students with and without SEN-L. Analyzing the patterns of correlations between SEN-L students' cognitive profiles and their test performance within and across experimental conditions and comparing these patterns between homogeneous subgroups of students with SEN-L will present us with additional information on the validity and formation of test achievements seen in students with SEN-L.

One of the aims of the feasibility studies described above is to examine the effects of adjusted test material and administration modes for the target population. Thus, despite some additional breaks the testing procedure has been set up to be as comparable as possible to the main samples', that is, the paper-and-pencil tests were administered in group settings and there was no extended testing time for students with SEN-L which was decided by the lack of consistent findings in using *extended time* as an accommodation.

Specifically, the first day of testing was largely consistent with testing procedures of 5th graders attending regular schools (see Table 2).

Table 2: The NEPS survey design in Grade 5 for the sample of students from regular schools of the wave 2010

Booklet 1	Booklet 2
Reading speed test	
Reading competence test	Mathematical competence test
Mathematical competence test	Reading competence test
Break	
Domain-general cognitive functioning (nonverbal)	
Orthography	
Break	
Student's questionnaire	

The additional diagnostic tests were administered on a second day of testing which took place about one week after the first.

To assure students with SEN-L's understanding of the given instructions and the correct appliance of tests, instruction demonstration posters have been used by test administrators in the testing sessions. Thus, the mode of information processing had been changed from just reading the instruction at the first pages of the test booklet by oneself to interacting with the test administrator while responding to the given examples. This approach was applied in the control group of students attending *Hauptschule* as well.

6. Conclusion

Including students with SEN into a nationwide, large-scale education study like the NEPS is not a straightforward endeavor. Nevertheless, NEPS is meeting this challenge by oversampling students with special educational needs in the area of learning within special schools, and by identifying sampling strategies for students with SEN-L attending regular school settings. These assessments will contribute substantially to exploring students with SEN-L in Germany. In the oversampling of students with SEN-L the data collected from parents, teachers, and school principals of students with SEN-L is comparable to the NEPS main studies and will be available for the scientific community via scientific use files for example. Thus, the NEPS contributes significantly to closing a research gap. With respect to the measurement of competencies, NEPS is attempting to balance the special needs of students with SEN-L on the one hand and the methodical and statistical requirements of gathering valid and comparable large-scale data on the other hand. So far, the feasibility studies outlined in this paper try to find perspectives to fulfill the demands on both sides. However, whether students with SEN-L can be included into the large-scale assessments of competencies and if so, how this can be achieved in a valid and comparable way will depend on empirical results. Apart from that, the feasibility studies will add to our knowledge on cognitive abilities of students with SEN-L and how to assess these in any case.

References

AERA – American Educational Research Association. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

Antor, G., & Bleidick, U. (Eds.). (2001). *Handlexikon der Behindertenpädagogik: Schlüsselbegriffe aus Theorie und Praxis*. Stuttgart, Germany: Kohlhammer.

Bleidick, U. (1968a). Über Beziehungen zwischen Milieuschädigung, Erziehungsschwierigkeit und Lernbehinderung. *Zeitschrift für Heilpädagogik, 19*, 225–240.

Bleidick, U. (1968b). Über Lernbehinderung: Begriffliche und psychodiagnostische Überlegungen. *Zeitschrift für Heilpädagogik, 19*, 449–464.

Blossfeld, H.-P., Maurice, J. von, & Schneider, T. (2011). The National Educational Panel Study: Need, main features, and research potential. 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. 5–17). Wiesbaden, Germany: VS. doi: 10.1007/s11618-011-0178-3

BMAS – Bundesministerium für Arbeit und Soziales. (Ed.). (2010). *Übereinkommen über die Rechte von Menschen mit Behinderungen*. Retrieved from http://www.bmas.de/SharedDocs/Downloads/DE/uebereinkommen-ueber-die-rechte-behinderter-menschen.pdf?__blob=publicationFile

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., Bonsen, M., & Gröhlich, C. (Eds.). (2009). *KESS 7: Kompetenzen und Einstellungen von Schülerinnen und Schülern der Jahrgangsstufe 7*. Münster, Germany: Waxmann.

Cormier, D., Altman, J., Shyyan, V., & Thurlow, M. D. C. (2010). *A summary of the research on the effects of test accommodations: 2007-2008: Technical Report 56*. Retrieved from http://www.cehd.umn.edu/NCEO/onlinepubs/Tech56/TechnicalReport56.pdf

Eberwein, H. (1997). Lernbehinderung – Faktum oder Konstrukt? *Zeitschrift für Heilpädagogik, 48*, 14–22.

Elliott, J., Thurlow, M., & Ysseldyke, J. (1996). *Assessment guidelines that maximize the participation of students with disabilities in large-scale assessments: Characteristics and considerations*. Retrieved from http://www.cehd.umn.edu/NCEO/onlinepubs/archive/Synthesis/synthesis25.html

Frahm, S., Goy, M., Kowalski, K., Sixt, M., Strietholt, R., Blatt, I., Bos, W., & Kanders, M. (2011). Transition and development from lower secondary to upper secondary 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. 217–232). doi: 10.1007/s11618-011-0188-1

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.

Gresch, C., Leuze, K., & Solga, H. (2011). *Integrationsschüler im Nationalen Bildungspanel (IntFös)*. Retrieved from http://www.wzb.eu/de/forschung/bildung-und-arbeit/bildungspanel-1

Hammill, D. D. (1990). On defining learning disabilities: An emerging consensus. *Journal of Learning Disabilities, 23*, 74–84. doi: 10.1177/002221949002300201

Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen (KFT 4-12+R)*. Göttingen, Germany: Beltz.

Hörmann, B. (2007). *Die Unsichtbaren in PISA, TIMSS & Co. Kinder mit Lernbehinderung in nationalen und internationalen Schulleistungsstudien*. Diploma thesis. University of Vienna, Austria. Retrieved from http://othes.univie.ac.at/86/1/Diplomarbeit.pdf

Institut für Schulentwicklungsforschung & Research School Education and Capabilities. (n.d.). *PARS. Eine Längsschnittstudie in NRW an weiterführenden Schulen*. Retrieved from http://www.bilderstrom.de/cms/fritz-winter-gesamtschule/fileadmin/fritz-winter-gesamtschule/medien/schulleben/pars/PARS_Presse_Info.pdf

Johns, J., & Vanleirsburg, P. (1992). Teaching test-wiseness: Can test scores of special populations be improved? *ERIC Research Review. Reading Psychology, 13*, 99–103. doi: 10.1080/0270271921301080

Jude, N., & Klieme, E. (2010). Das Programme for International Student Assessment (PISA). 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. 11–21). Münster, Germany: Waxmann.

Kanter, G. O. (1997). Lernbehindertenpädagogik – Gegenstandsbestimmung, Begriffsklärung, sowie: Lernbehinderungen und die Personengruppe der Lernbehinderten. In G. O. Kanter & O. Speck (Eds.), *Pädagogik der Lernbehinderten: Vol. 4. Handbuch der Sonderpädagogik* (pp. 7–64). Berlin, Germany: Marhold.

Kanter, G. O. (2001). Lernbehinderung, Lernbehinderte, Lernbehindertenpädagogik. In G. Antor & U. Bleidick (Eds.), *Handlexikon der Behindertenpädagogik. Schlüsselbegriffe aus Theorie und Praxis* (pp. 119–124). Stuttgart, Germany: Kohlhammer.

KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (1994). *Empfehlungen zur sonderpädagogischen Förderung in den Schulen in der Bundesrepublik Deutschland.*

Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1994/1994_05_06-Empfehl-Sonderpaedagogische-Foerderung.pdf

KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2009). *Das Bildungswesen in der Bundesrepublik Deutschland 2008. Darstellung der Kompetenzen und Strukturen sowie der bildungspolitischen Entwicklungen für den Informationsaustausch in Europa*. Deutsche EURYDICE-Informationsstelle der Länder im Sekretariat der Kultusministerkonferenz in Zusammenarbeit mit der Deutschen EURYDICE-Informationsstelle des Bundes beim BMBF. Retrieved from http://www.bildungsserver.de/glossar.html

KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2012). *Sonderpädagogische Förderung in Schulen 2001 bis 2010*. Retrieved from http://www.kmk.org/fileadmin/pdf/Statistik/Dokumentation_SoPaeFoe_2010.pdf

Koretz, D., & Barton, K. (2003). *Assessing students with disabilities: Issues and evidence: CSE Technical Report 587*. Retrieved from http://www.cse.ucla.edu/products/reports/TR587.pdf

Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens* (2nd ed.). Vienna, Austria: Hogrefe.

Lauth, G. W., Brunstein, J. C., & Grünke, M. (2004). Lernstörungen im Überblick: Arten, Klassifikation, Verbreitung, Entwicklungsperspektiven. In G. W. Lauth, M. Grünke, & J. Brunstein (Eds.), *Interventionen bei Lernstörungen* (pp. 13–23). Göttingen, Germany: Hogrefe.

Lehmann, R. H., & Hoffmann, E. (Eds.). (2009). *BELLA. Berliner Erhebung arbeitsrelevanter Basiskompetenzen von Schülerinnen und Schülern mit Förderbedarf „Lernen“*. Münster, Germany: Waxmann.

Lehmann, R. H., Nikolova, R., & Peek, R. (2004). *Lernausgangslage arbeitsrelevanter Basiskompetenzen der Berliner Schülerinnen und Schüler mit dem Förderschwerpunkt Lernen in den Jahrgängen 8-10 (LABEL 8-10) – Erste Befunde*. Berlin, Germany: Senatsverwaltung für Bildung, Jugend und Sport.

Lehmann, R. H., & Peek, R. (1997). *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen: Bericht über die Untersuchung im September 1996 (unpublished technical report)*. Hamburg, Germany.

McDonnell, L. M., McLaughlin, M., & Morrison, P. (Eds.). (1997). *Educating one and all: Students with disabilities and standards-based reform*. Washington, DC: National Academy Press.

Meyerhöfer, W. (2007). Testfähigkeit – Was ist das? In S. T. Hopmann, G. Brinek, & M. Retzl (Eds.), *PISA zufolge PISA – PISA according to PISA*. Vienna, Austria: LIT-Verlag.

Moosbrugger, H., & Oehlschlägel, J. (1996). *Frankfurter Aufmerksamkeits-Inventar: FAIR; Testmanual* (1st ed.). Bern, Switzerland: Huber.

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

OECD – Organisation for Economic Co-operation and Development. (2006). *PISA assessing scientific, reading and mathematical literacy. A framework for PISA 2006*. Paris, France: OECD.

Prenzel, M., Carstensen, C. H., Frey, A., Drechsel, B., & Rönnebeck, S. (2007). PISA 2006 – Eine Einführung in die Studie. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme, & R. Pekrun (Eds.). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (pp. 31–59). Münster, Germany: Waxmann.

Prenzel, M., Drechsel, B., Carstensen, C. H., & Ramm, G. (2004). PISA 2003 – eine Einführung. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H. G. Rolff, J. Rost, & U. Schiefele (Eds.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (pp. 13–46). Münster, Germany: Waxmann.

Schmetz, D. (1999). Förderschwerpunkt Lernen. *Zeitschrift für Heilpädagogik, 4*, 134–143.

Schröder, U. (2000). *Lernbehindertenpädagogik: Grundlagen und Perspektiven sonderpädagogischer Lernhilfe*. Stuttgart, Germany: 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

Stechow, E. von (2006). PISA und die Folgen für schwache Schülerinnen und Schüler. *Vierteljahreschrift für Heilpädagogik und ihre Nachbargebiete, 75*, 285–292.

US Department of Education. (2004) *Individuals with disabilities education act*. Public Law 108–446, 108th Congress.

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Göttingen, Germany: Hogrefe.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C. H. (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.

Weiß, R. H. (1978). *Grundintelligenztest Skala 2 – CFT 20: Handanweisung für die Durchführung, Auswertung und Interpretation*. Braunschweig, Germany: Westermann.

Werning, R., & Löser, J. M. (2010). Inklusion: aktuelle Diskussionslinien, Widersprüche und Perspektiven. *DDS – Die Deutsche Schule. Zeitschrift für Erziehungswissenschaft, Bildungspolitik und pädagogische Praxis, 102*, 103–114.

Wocken, H. (2000). Leistung, Intelligenz und Soziallage von Schülern mit Lernbehinderungen: Vergleichende Untersuchungen an Förderschulen in Hamburg. *Zeitschrift für Heilpädagogik, 51*, 492–503.

Wocken, H. (2005). *Andere Länder, andere Schüler? Vergleichende Untersuchungen von Förderschülern in den Bundesländern Brandenburg, Niedersachsen und Hamburg. Forschungsbericht*. Retrieved from http://www.mbjs.brandenburg.de/sixcms/media.php/5527/wocken_ergebnis-heft.pdf

Wocken, H., & Gröhlich, C. (2009) Kompetenzen von Schülerinnen und Schülern an Hamburger Förderschulen. In W. Bos & M. Bonsen (Eds.), *KESS 7: Kompetenzen und Einstellungen von Schülerinnen und Schülern der Jahrgangsstufe 7* (pp. 133–142). Münster, Germany: Waxmann.

Zielinski, W. (1995). *Lernschwierigkeiten: Ursachen – Diagnostik – Intervention*. (2nd ed.). Stuttgart, Germany: Kohlhammer.