14

Schriften aus der Fakultät Humanwissenschaften
der Otto-Friedrich-Universität Bamberg

The Development of Reading Literacy from Early Childhood to Adolescence

Empirical Findings from the Bamberg BiKS Longitudinal Studies

Maximilian Pfost, Cordula Artelt, Sabine Weinert (Eds.)

University of Bamberg Press

14 Schriften aus der Fakultät Humanwissenschaften der Otto-Friedrich-Universität Bamberg

Schriften der Fakultät Humanwissenschaften
der Otto-Friedrich-Universität Bamberg

Band 14

University
of Bamberg
Press
2013

The Development of Reading Literacy from Early Childhood to Adolescence

Empirical Findings from the Bamberg BiKS Longitudinal Studies

Edited by
Maximilian Pfost, Cordula Artelt, and Sabine Weinert

University of Bamberg Press
2013

Bibliographische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Informationen sind im Internet über http://dnb.ddb.de/ abrufbar

Dieses Werk ist als freie Onlineversion über den Hochschulschriften-Server (OPUS; http://www.opus-bayern.de/uni-bamberg/) der Universitätsbibliothek Bamberg erreichbar. Kopien und Ausdrucke dürfen nur zum privaten und sonstigen eigenen Gebrauch angefertigt werden.

Herstellung und Druck: docupoint GmbH, Barleben
Umschlaggestaltung: University of Bamberg Press, Andra Brandhofer
Englisches Korrektorat: Dr. Jane Zagorski
Text-Formatierung: text & kontext, Bamberg

© University of Bamberg Press Bamberg 2013
http://www.uni-bamberg.de/ubp/

ISSN: 1866-8674
ISBN: 978-3-86309-164-4 (Druckausgabe)
eISBN: 978-3-86309-165-1 (Online-Ausgabe)
URN: urn:nbn:de:bvb:473-opus4-37708

Contents

Chapter 1

1 Editorial[1]

Maximilian Pfost, Cordula Artelt, and Sabine Weinert

It is important to handle written information efficiently for participating in economic, cultural, and social life of modern societies (OECD, 2003). Text is omnipresent at all niches of life. Even when using the public transport system, we need to be able to read the fares, to handle the vending machine, and to identify the right bus line in order to reach our goal. There is general agreement that, at least up to a basic level, being able to read is essential for life, not just for the individual but also for the well-being of the whole society (UNESCO, 2005). In order to maximize individual life chances, every child should be given the possibility to learn to read and to be able and motivated to use this skill effectively and on a high level. For providing such learning opportunities, researchers as well as educators need to understand how individuals acquire the ability to read and why some learn and practice it so successfully whereas others struggle or fail.

Author Note

Maximilian Pfost,
Department of Educational Research, University of Bamberg, Germany.

Cordula Artelt,
Department of Educational Research, University of Bamberg, Germany.

Sabine Weinert,
Department of Developmental Psychology, University of Bamberg, Germany.

Correspondence concerning this chapter should be addressed to Maximilian Pfost, Department of Educational Research, University of Bamberg, Markusplatz 3, 96045 Bamberg, Germany. E-Mail: maximilian.pfost@uni-bamberg.de

[1] The studies were supported by grants of the German Research Foundation (DFG) to the Research Group “BiKS” (“Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Grundschulalter”; English: “educational processes, competence development, and selection decisions in preschool and school-age children”) at the Otto-Friedrich-Universität Bamberg, Germany (principle investigators: Cordula Artelt, Peter Blossfeld, Gabriele Faust, Hans-Guenther Rossbach and Sabine Weinert).

Reading and understanding written information is a complex process that goes far beyond the ability of a simple recoding of letters. Reading comprises processes that range from the decoding of letters and word recognition to activities related to the comprehension of words, phrases, and paragraphs including the regulation of such activities (see Snowling & Hulme, 2005, for a comprehensive review). The studies that are reported within this volume analyze student's reading literacy development and its precursors and predictors in different critical developmental periods that range from early preschool years up to secondary school. Empirical research in general has accumulated evidence of high mean rates of improvement in literacy in the course of this developmental period in combination with an overall trend of declining growth rates as students become older (Bloom, Hill, Black, & Lipsey, 2008; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996). According to Hill, Bloom, Black, and Lipsey (2008), annual growth rates for reading vary between one and a half standard deviation at the beginning of primary school and almost monotonically decline up to less than a tenth standard deviation at the end of secondary school. The studies assorted in this book thereby focus on analyzing individual differences in these reading literacy trends. Until to date, individual differences in reading competencies have been well studied using cross-sectional datasets like PIRLS (Bos, et al., 2007; Mullis, Martin, Kennedy, & Foy, 2007) or PISA (Baumert, et al., 2001; Klieme, et al., 2010; OECD, 2010). However, there is much less empirical research analyzing individual differences in reading literacy using longitudinal datasets and with a focus on developmental changes.

The BiKS-longitudinal and multi-cohort study provides excellent conditions for researchers interested in such developmental questions and who try to better understand the complex network of factors influencing students' cognitive development. BiKS is the German acronym for *"Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Schulalter"* which might be best translated as "Educational processes, competence development, and selection decisions in preschool- and school age". A description of the BiKS-longitudinal studies, including some of its' major goals and perspectives, is provided in the second chapter of this volume. The authors of the second chapter, Christian Lorenz, Monja Schmitt, Simone Lehrl, Michael Mudiappa, and Hans-Guenther Rossbach furthermore provide background information that led to the decision of creating an interdisciplinary research group for longitudinal research in the field of education. Finally, an overview of the two individual studies, their organizational structure, and the characteristics of each sample is depicted.

The three chapters that follow are devoted to questions concerning the development of reading literacy between preschool and primary school. Every chapter focuses on a different agent that influences children's individual early literacy development. Specific to these chapters is their multimethodological approach, relating observational data to questionnaire and test data. The opening is made by Susanne Lehrl, Susanne Ebert, and Hans-Guenther Rossbach (Chapter 3). In their chapter, the role of the family for reading literacy development is highlighted. The authors examine the influence of specific home literacy practices for children of preschool age, like shared book reading or the teaching of literacy, on the development of children's basic reading skills and their reading comprehension in primary school. In their ambitious study, the authors combine self-reported questionnaire data of the parents with behaviour observations of parent-child-interactions and objective test data of the students before and after the transition from preschool to primary school. The authors show that different facets of the home learning environment are important for student's basic reading skills and reading comprehension. Furthermore, the mediating role of emergent literacy skills is highlighted.

The subsequent chapter by Susanne Kuger, Hans-Guenther Rossbach, and Sabine Weinert (Chapter 4) focuses on the role of preschools in the development of children's reading literacy. In this chapter, the authors investigate the relation between differences in the quality of classroom stimulation as a whole on the one hand and stimulation quality experienced by the individual child on the other hand with the development of children's reading literacy. Surprisingly, there seems to be only little emphasis in German preschools on fostering early literacy skills. Furthermore, the observed differences in preschool activities supporting code related skills do not prove to be important for student's later reading comprehension whereas more general facets of literacy and language support enhance student's later reading competence.

In Chapter 5 in contrast, Susanne Ebert and Sabine Weinert focus on how children's language competencies in early preschool age impact the development of reading literacy four years later. As language is multi-componential in its nature, the specific importance of different facets of language for different aspects of reading literacy is considered in this chapter. The author's results underline the specific share of (a) phonological information processing skills on later basic reading skills, and (b) the importance of linguistic abilities (vocabulary and grammar) for reading comprehension in the second year of primary school when controlling for basic reading skills that may hinder more complex comprehension processes from taking place. Interestingly, (c) integrative language competencies (such as story reproduction and comprehension as well as sentence reproduction) in preschool did not impact later reading literacy over and above the impact of language components (vocabulary, grammar and phonological skills). The results are discussed within the broader debate of how and in which ways language skills are related to reading literacy development.

The second empirical section, comprising Chapters 6 to 9, focuses on the development of children's reading literacy during the transition from primary to secondary school. Analyses of these chapters are based on the second, older cohort of the BiKS-longitudinal studies. First, in chapter 6, Thorsten Schneider and Maximilian Pfost are tracing social disparities in literacy development of students from families with and without immigration background. Thereby, the role of cultural capital and cultural activities within families as a mechanism for the development of these differences is investigated. Results indicate an increasing achievement gap between students of families with different educational background. However, there is a tendency that this effect is more pronounced for students of native families than for students of families with an immigration background. The findings are related to the debate of whether and to what extend cultural resources are transferable between countries and social-cultural contexts.

In the next chapter, Irene Schurtz, Tobias Dörfler, Maximilian Pfost, and Cordula Artelt (Chapter 7) address the development of students' interest in language arts and its relations to the development of reading literacy in secondary school. Because motivation is one of the key components that is used to explain individual differences in reading, this chapter tries to relate the concept of interest with measures of actual reading behaviour and the development of reading literacy. The authors confirm their expectation of a general negative developmental trend for interest in language arts in secondary school. Furthermore, only weak relations of students' interest in language arts and students' reading competence are shown whereas more close relations to students' reading activities are prevalent.

The subsequent two chapters of this volume are dedicated to the role of school for reading literacy development. First, in chapter 8, Maximilian Pfost and Cordula Artelt ask whether attending different types of schools is related to differences in the development of reading literacy. Thereby developmental trajectories of students attending different types of schools between Grade 5 and Grade 7 are traced. In a second part of their study, the effect of attending the upper academic track in comparison to attending the lower and middle academic track is estimated. In their analyses, the authors try to determine effects of attending different school tracks independent of the student's individual characteristics. According to their results, increasing competence differences between the different school tracks are shown for measures of reading comprehension but not for vocabulary. Furthermore, different learning environments that go along with the school tracks contribute to this fan-spread effect.

In Chapter 9 finally, Constance Karing, Maximilian Pfost, and Cordula Artelt concentrate on the diagnostic competence of teachers in the domain of reading and ask for its consequences for the development of students' reading literacy. The authors demonstrate empirically that teachers' diagnostic competence is positively related to the development of students' reading competence. Furthermore, this relation is moderated by instructional variables such as the degree of individualization of lessons.

In summary, this volume provides convincing empirical evidence for the importance of a view that learning to read is not limited to experiences made in schools. Schools are of special importance, but further institutions such as preschools influence the acquisition of reading related skills just as well as further variables beyond the formal education system. The family and parents of each student for example are one of these sources contributing to success or failure in learning to read. Across studies, the findings of the BiKS-longitudinal study have shown that individual differences in reading literacy arise due to schools and preschools, teachers and educators in school and preschool, parents as well as the student's own cognitive and conative characteristics. In addition, we need to keep in mind that such influences, although they were treated separately in the presented analyses, are interacting with each other.

The identification and description of these variables, as has been done by the presented studies, however provides further support that in order to better understand reading literacy, longitudinal empirical research covering several years of individual development is needed.

References

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, J. M., et al. (Eds.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1,* 289-328. doi:10.1080/19345740802400072

Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., et al. (Eds.). (2007). *IGLU 2006. Lesekompetenz von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: a longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88,* 3-17. doi:10.1037/0022-0663.88.1.3

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2,* 172-177. doi:10.1111/j.1750-8606.2008.00061.x

Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., et al. (Eds.). (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.

Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 International report*. Boston, MA: TIMSS & PIRLS International Study Center.

OECD. (2003). *Literacy Skills for the World of Tomorrow - Further Results from PISA 2000.* PISA: OECD Publishing. Retrieved from http://www.oecd.org/pisa/pisaproducts/pisa2000/literacyskillsfortheworldoftomorrowfurtherresultsfrompisa2000-publications2000.htm

OECD. (2010). *PISA 2009 Results: Learning to Learn – Student Engagement, Strategies and Practices (Volume III)*. PISA: OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/education/pisa-2009-results-learning-to-learn_9789264083943-en

Snowling, M. J., & Hulme, C. (Eds.). (2005). *The science of reading: A handbook*. Malden, MA: Blackwell Publishing.

UNESCO. (2005). *Education for All Global Monitoring Report: Literacy for life*. Paris: UNESCO Publishing. Retrieved from http://unesdoc.unesco.org/images/0014/001416/141639e.pdf

Chapter 2

2 The Bamberg BiKS Research Group[1]

Christian Lorenz, Monja Schmitt, Simone Lehrl, Michael Mudiappa, and Hans-Guenther Rossbach

Summary

BiKS, the German acronym for the current study, stands for “educational processes, competence development, and selection decisions in preschool and school-age children.” The present chapter provides an overview of the research conducted within in the research group, the study’s design, its samples, participants, and assessments. The interdisciplinary research group was supported by the German Research Foundation (DFG) and conducted by several researchers from psychology, education, and sociology. Across a period of more than 7 years, the study followed more than 4,000 Bavarian and Hessian children in two cohorts

Author Note

Christian Lorenz,
National Educational Panel Study, University of Bamberg, Germany;

Monja Schmitt,
Institute of Sociology, University of Bamberg, Germany;

Simone Lehrl,
Early Childhood Education, University of Bamberg, Germany;

Michael Mudiappa,
National Educational Panel Study, University of Bamberg, Germany;

Hans-Guenther Rossbach,
Early Childhood Education, University of Bamberg, Germany.

Correspondence concerning this chapter should be addressed to Christian Lorenz, National Educational Panel Study, University of Bamberg, 96045 Bamberg, Germany.
E-mail: christian.lorenz@uni-bamberg.de

[1] The research group was supported by the German Research Foundation (grants to C. Artelt, P. Blossfeld, G. Faust, H.-G. Roßbach, S. Weinert, and colleagues)

across very decisive phases of their academic careers. The first cohort, called BiKS-3-10, focused on 547 children from the age of 3 when they had just entered preschool until the end of primary school in Grade 4 and followed an additional 443 children attending the same classes across primary school. In the second sample, BiKS-8-14, a total of 2,395 students were assessed during the same time period from Grade 3 across the transition to secondary school until the end of secondary school in Grade 9. After the transition into secondary school, the sample was augmented by an additional 879 secondary school students. Not only the children, but their families, their preschool teachers, and their teachers were involved in the study as well.

Objectives of the study

As international studies on student assessment have shown, there are serious deficits in the German school system with regard to the students' achievement (Baumert et al., 2001). After the so-called "PISA shock" of the year 2000, the achievement of German students improved overall as shown in recent PISA assessments (OECD, 2010).

It is widely accepted that the development of the students' achievement is a result of their predispositions as well as their cumulative experiences in academic, pre-academic, and family contexts (e.g., Baumert et al., 2001; Hattie, 2009), but details about the factors that have contributed to the (differential) development of student achievements are still needed. This was one of the reasons why the BiKS research group (the German acronym for "Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Grundschulalter"; English: "educational processes, competence development, and selection decisions in preschool and school-age children") was founded in 2005 at the Otto-Friedrich-University in Bamberg, Germany. The BiKS research group is supported by the German Research Foundation (DFG) to provide a close cooperation between researchers from psychology, education, and sociology to study the diverse factors that contribute to children's development. From a longitudinal perspective, BiKS focuses on developmental processes that are relevant to education and achievement in preschool as well as in elementary and secondary school by studying children from the ages of 3 to 15 in two panel studies that are aligned with each other.

BiKS also looks closely at the formation of transition decisions, examines their preconditions, and follows the effects of these decisions across the ensuing school years. Within that time period, the transitions from preschool institution[2] to elementary school and – especially in the German multi-tracked school system – from elementary to secondary school constitute important milestones for the children and for the research as well. In addition to the (pre)conditions in which the students live and learn within their institutional and family contexts, the relationships between these contexts are also important within the BiKS project.

The subprojects within the BiKS longitudinal study

When the study was first designed, one of the main intentions was to create a close research network of representatives from education, psychology, and sociology. To address the multifaceted research questions of the study, the BiKS project is divided into eight subprojects with different foci with regard to their main research topic. Each of these subprojects, which will be introduced below, consists of several researchers who work autonomously but within the framework of the overall project and belong to one of the previously mentioned disciplines. However, although all subprojects will be introduced, Subprojects 2, 3, 4, and 8 (listed below) are of special importance with regard to this book as these focus on facets of reading development.

Subproject 1

"Framing project – familial and institutional conditions for the linguistic and cognitive development of children's abilities and decisions concerning children's education in preschool and school-age children (longitudinal studies)"

Subproject 1 is responsible for the coordination of all subprojects and the supervision of the surveys run by the BiKS research group (see section "Design of the study"). The investigation of relations and interdependencies between the development of the children's abilities and educational decisions are based on two longitudinal studies. In

---

[2] By this we mean the German 'Kindergarten', a pre-school establishment for children aged between three and six as part of child and youth welfare services - may be either publicly or privately maintained [not part of the school system].

these studies, children, their parents, and their preschool, primary school, and secondary school teachers are subjects of periodic research across a longer period. Several tests, questionnaires, interviews, and observations are used to examine the children's developing abilities as well as the parents', preschool teacher's and teacher's assessments, and more.

Subproject 2

"Longitudinal effects of the global and domain-specific quality of stimulation in the family, preschool, and elementary school on early childhood competence development"

Main focuses of Subproject 2 are the different learning environments the child is engaged in from early to middle childhood, how to measure quality of these learning environments and their effects on the development of language and more general cognitive competencies throughout early and middle childhood. In detail, the central aims of Subproject 2 are to investigate the effects of structural conditions, educational beliefs and domain-specific processes in early family and preschool settings on early childhood development and, as children go on to elementary school, the additional and interactional effects of the next institutional setting in the course of the children's development. Another aspect of Subproject 2 extends the longitudinal section to the last 2 years of elementary school to examine the characteristics that are important to this period of time in their educational trajectories (e.g., changes in curriculum, transition to secondary school, more peer contacts).

Subproject 3

"Analysis of the relation between language acquisition, (meta)cognitive development, and characteristics of adult-child interactions"

In the context of the overall study, Subproject 3 is responsible for the selection, development, and testing of instruments for measuring the abilities and skills of the children participating in the BiKS-3-10 sample. These measures include indicators of domain-specific as well as domain-general aspects of individual development. In addition to various measures of language and cognitive development selected control variables such as motivational aspects, self-concept, and personality variables are assessed.

Additionally, Subproject 3 is concerned with the analysis of the relation between language acquisition, cognitive development, and metacognitive progress (i.e., the acquisition of knowledge about knowledge and thought processes including children's developing "theory of mind"). A special interest lies in developmental differences in language and (meta)cognitive development due to social disparities. To investigate the impact of learning environment on these developments in more detail videos of adult-child interactions are analyzed. These include parent-child interaction situations (play, picture-book reading) at preschool age and teacher-child situations (classroom observation) in school age. Indicators derived from these interaction situations supplement measures assessed in Subproject 2. A special focus is on the developing (academic) language competencies of children, influencing variables and predictors (including characteristics of teachers' language) and their impact on school performance.

Subproject 4

"The development of students' competencies and interests in primary and secondary school"

Subproject 4 investigates the development of students' school competence development in the domains of mathematics, reading, and English as a foreign language and tries to explain interindividual differences by factors that occur at the school, classroom, and individual levels. In addition to the question of the interindividual stability of students' competence development, differential pathways of students' competencies for different groups of students, (e.g., different socio-economic backgrounds or different scholastic promotion) are demonstrated and linked to possible mediating processes. The second focus of the subproject is on the analysis of processes involved in the differentiation of students' interests. With respect to individual competence levels and subjective competence beliefs, we ask whether the expected decrease in the mean interest level can be attributed to processes of internal differentiation in favor of certain domains or subjects. Finally, the project focuses on the analysis of interrelations between competence and interest development and asks for instructional conditions that can promote successful development in both domains.

Subproject 5

"The formation of educational decisions in primary and secondary school"

Subproject 5 deals with educational decisions in primary school and during secondary school. In the first phase of the project, the focus was on the transition from primary to secondary school. After that, the stabilization or revision of the school choice was researched. Currently, the subproject plans to analyze the formation of educational decisions or those concerning vocational training at the end of lower secondary school in detail. With this new focus, the subproject works on five specific subjects:

1) The reconstruction of educational pathways to explain the differences of the cross-sectional results between PISA and IGLU,
2) the influence of institutional differences in frame conditions of secondary school on individual decision options,
3) the meaning of different actors (parents vs. peers) for educational decisions,
4) the influence of contextual conditions of school and non-school related educational processes and training facilities on decisions concerning education and vocational training respectively, and
5) the analysis of the development of school or work related interests and educational aspirations.

Subproject 6

"Formation of decision-making processes in connection with expectations in education and the development of competence: Transitions into primary school"

The central aspects of Subproject 6 are the ways in which parents and educators deal with primary school, the educational institution that follows kindergarten. A key phrase of the study is "school-readiness." The project analyzes parents' and educators' understanding of this phrase and whether and how they assist their child's development in this respect. Parents have limited input with regard to the age at which their children move on to primary school as well as the choice of the school itself. The study asks about the parents' preferences for an earlier or later transition into primary school and about the point in time at which these questions become important to the parents. Which views do parents with a Turkish immigration background and their children's teachers hold? In Bavaria in particular, those questions are of special interest because the school system intends to change the due dates concerning the age for mandatory school attendance for children, and this will result in younger primary school students there. Subproject 6 examines how the final decision about the child's schooling arrangements is formed based on the combined views and plans of the parents, the kindergarten and the chosen primary school. In addition, the success of school enrolment is observed in the view of parents and children and with regard to academic skills. Subproject 6 works in collaboration with Subprojects 2 and 5 by including corresponding questions in the surveys with parents and educators. Open guideline interviews were held with a small group of parents – including Turkish-speaking parents from Bavaria and Hesse.

Subproject 7

"Competence development and educational decisions of immigrant children in primary and secondary school"

Subproject 7 investigates the competence development of students with immigration backgrounds and parental decisions regarding the educational careers of their children in primary and lower secondary school. The research questions of this subproject address the educational aspirations of parents which are immigrants, the differences between parents and teenagers with and without an immigration status regarding the revision or stabilization of educational decisions, and the development of the occupational and educational aspirations of teenagers with immigration backgrounds. Furthermore, the perception of discrimination and gender-specific disparities in connection with school performance and aspirations are considered. Subproject 7 therefore analyzes quantitative data and conducts qualitative interviews with Turkish immigrant parents and their children.

Subproject 8

"Prerequisites, structure, and effects of teachers' diagnostic competence"

Subproject 8 focuses on the structure, the prerequisites, and the effects of teachers' diagnostic competence. This project aims to investigate the accuracy of teachers' diagnostic judgments concerning students' competencies in three different school-related subject domains (German, Mathematics, and English) as well as judgments about students' motivations and emotions. The major goal of this subproject is to analyze the precursors and prerequisites of diagnostic competence. To this end, we differentiate between features of the class, the judgment object, and the teacher as predictor variables. In addition, the study assesses which of these variables might mediate the effect of teachers' diagnostic competence on students' performance.

Within the context of an additional study, the professional knowledge base of teachers' diagnostic competence was investigated in the domain of text comprehension. Thus, we were interested in the teachers' knowledge about factors affecting the difficulty of tasks and text characteristics and text comprehension strategies. Moreover, the variability and the promotion of the knowledge base were examined by comparing teachers with different professional backgrounds.

Design of the study

The BiKS research group runs a two-cohort longitudinal study using two different samples that are linked to each other in several ways. Both studies were originally designed to run for seven years.

In the BiKS-3-10 longitudinal study, the development of children's abilities, the influence of home learning environment and preschool quality, and decisions concerning the children's education – especially regarding the transition from preschool to primary school – are the objects of investigation. Beginning in the fall of 2005, an initial group of 547 3- and 4-year-old children were observed from the time they entered preschool across a period of 7 years until they had finished the fourth grade of primary school.

In the BiKS-8-14 longitudinal study, the objects of investigation are the development of children's abilities and decisions concerning the children's education – especially with regard to the development of the children's marks, competencies, interests, and aspirations as well as the transition from primary school to secondary school and the results of the decision to place a child in a special track. Beginning in the spring of 2006, a group of 2,395 primary school students were followed from the beginning of the third grade across a period of 6 years until they completed the ninth grade of secondary school.

Many resources were dedicated to drawing the two samples to guarantee the quality of the research. Therefore, several criteria were formulated to guide the sampling procedures of both studies. The first criterion was a stratification by federal state. Sixty percent of the participants stemmed from Bavaria where the BiKS study is native, and 40% came from Hesse. The two federal states have differences with regard to their educational policies and institutional conditions, among others. A second stratification occurred with respect to city size. One third of the participants lived in major cities (Frankfurt/M. and Nuremberg); the other two thirds lived in market towns and rural regions. Furthermore, facilities (i.e., schools and kindergartens) with low, medium, and high immigration ratios each provided one third of the children. Last, an equal distribution of the number of groups per facility (1 to 3) was attempted.

The bar graph shown in Figure 1 illustrates that the BiKS measurement points cover an age range from kindergarten to the end of grade 9, with a 1.5 year overlap in grade 3 and 4.

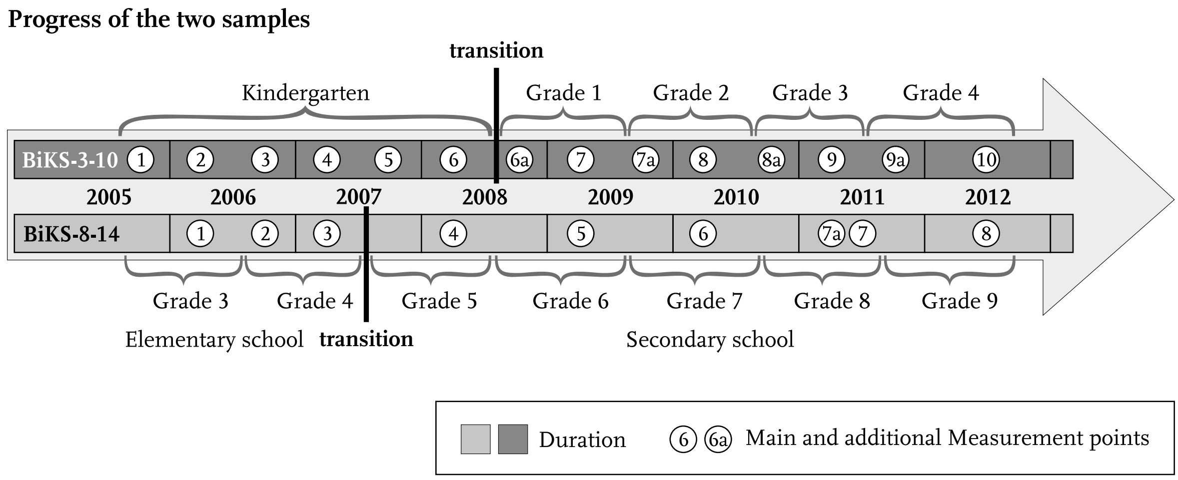

Figure 1. Time bar with measurement points of the two BiKS samples.

In the following section, the design of the two studies will be described in more detail, including the development of the samples from the beginning of the BiKS longitudinal study until now as well as the wide variety of measurement instruments that have been applied across the years.

Design and study participants: BiKS-3-10

The BiKS-3-10 study uses a stratified (by immigration status, region, and federal state) random sample to survey the effects of different contexts on the processes that are relevant to the development and fostering of the children as well as to survey the effects of tracking decisions that are made in the school system. For better comparability across kindergartens, special facilities such as outdoor or integrative kindergartens and open facilities without regular groups were not included in the sample.

Table 1. Selected Characteristics of the BiKS-3-10 Sample at the Beginning of the Study

Sample size	N = 547 attending 97 Kindergarten groups
Children's mean age (t1)	M = 44.5 months (SD = 5.0)
Gender distribution	52.2% male 47.8% female
Family status	84.6% married 10.8% unmarried 4.4% divorced 0.2% widowed
Number of children in the families	23.0% families with a single child 51.7% families with two children 25.3% families with more than two children
Immigration background (by the parents' birth country)	78.1% no immigration background 11.7% one parent born in a foreign country 10.2% both parents born in foreign countries
Highest school leaving certificate in the family	0.7% no certificate 18.6% certificate of secondary education 31.3% general certificate of secondary education 46.3% general qualification for university entrance 3.1% foreign certificate

The original sample was recruited from 60 Bavarian and 37 Hessian preschool classes with 547 children in 97 kindergartens with a mean age of 44.5 months (cf. Table 1). The average number of children assessed per preschool class was 5.6. This number is not equivalent to the class size of the preschools as preschool classes were usually comprised of age-mixed groups and not all the children in a class did necessarily meet the inclusion criterion. Children were included in the study if mandatory school enrollment was due in the fall of 2008. Males comprised 52.2% of the children who were chosen to be in the sample. With respect to the initial sample and based on the parents' birth country, three fourths of the children (78%) were German, 12% had a mother or father who was born in a foreign country, and 10% of the sample had parents who were both immigrants. Nearly 8% of the children in the sample usually spoke a language other than German with their families. Furthermore, 85% of the parents were married, and in almost one half of the sample, the highest level of education in the families was the general qualification for university entrance.

In 2008, there were still 94 daycare centers participating in the study, corresponding to exactly 5 children per facility. However, this is only the number of children whose parents permitted them to participate. At each measurement point, part of the sample did not fill out some measures because they were absent for some reason; thus, the real level of participation was – depending on the measurement point and the instrument – approximately 2% to 8% lower than Figure 2 and Figure 3 suggest. In 2009, most children in the sample transitioned to elementary school. New classmates were asked to join the study, first, so that we could obtain more information about the class context with regard to the mean socio-economic status and the achievement level among other things, and second, to increase the number of children who began school 1 year before or 1 year after the usual point in time. When the children transitioned to elementary school at the expected age, 471 children (86% of the original sample) continued to participate in the BiKS study. We were then able to recruit an additional 528 families to participate. Thus, the sample size was increased to 999 children. In 2011, when most children in the sample were in the third grade, it was necessary to again ask the parents for their permission. Unfortunately, a substantial number of parents refused to agree to the further participation of their children so that the sample was reduced. (cf. Figure 2).

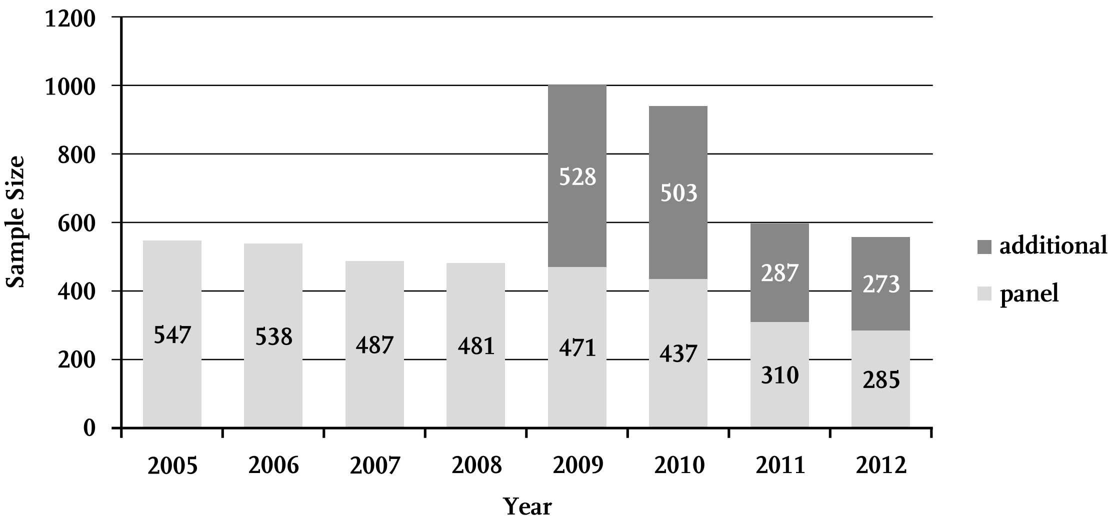

Figure 2. Sample enhancement and panel attrition for BiKS-3-10[3]

Of course, the children did not all begin school at the same time. It was especially challenging to longitudinally follow the children who began school one year earlier or one year later than the majority of their peers who began school at the expected age. An additional focus within the BiKS study lies on these children, but because they are not relevant to this book, they are merely mentioned here.

Design and study participants: BiKS-8-14

The second sample, BiKS 8-14, was initially recruited in 2006 with a total of 2,395 children who attended the third grade in 155 different classes distributed across 82 Bavarian and Hessian elementary schools. Their mean age was 9 years and 3 months; 52.2% were male. The sample was deliberately chosen from schools into which the children of the BiKS-3-10 sample would probably move after kindergarten. Thus, we were able to directly compare the measures and facets of the BiKS-3-10 sample with the BiKS-8-14 sample in the third and fourth grades with a temporal distance of 5 years in the same institutional context.

[3] Besides the main measurement points drawn in Figure 2, additional studies took place between them using subsamples for special research questions, the details of which cannot be given here.

Table 2. Selected Characteristics of the BiKS-8-14 Sample at the Beginning of the Study

Sample size	N = 2,395 attending 155 classes in 82 schools
Children's mean age (t1)	M = 111.1 months (SD = 5.7)
Gender distribution	52.2 % male 47.8 % female
Family status	83.6 % married 5.4 % unmarried 10.2 % divorced 0.8 % widowed
Number of children in the families	15.3 % families with a single child 51.5 % families with two children 33.2 % families with more than two children
Immigration background (by birth country of the parents)	73.5 % no immigration background 12.8 % one parent born in a foreign country 13.6 % both parents born in foreign countries
Highest school leaving certificate in the family	2.8% no certificate 22.0% certificate of secondary education 32.4% general certificate of secondary education 42.5% general qualification for university entrance 0.3% other

The sample characteristics were similar to the BiKS-3-10 sample. Slight differences existed with regard to family status, for which the proportion of divorced parents was more than twice as high, which was probably due to the higher age of the parents. Most likely for the same reason, the proportion of families with more than one child was somewhat higher in this older sample. The distribution of immigrants in the BiKS-8-14 sample was very similar to the one found in BiKS-3-10, but the percentage of children who usually did not speak German in their families was only half as high (i.e., 4%) as in the other sample.

After three measurement points, the children of the BiKS-8-14 sample moved from elementary school into secondary school. Then, for economic reasons, different approaches were used to follow the existing sample and enhance it with additional students from the classes the children moved into. The first approach that we applied affected about 800 children who could not be followed in the school context after they transitioned to secondary school (e.g., because they moved to schools outside of the research area). These children remained in the study by answering questionnaires that were sent by mail but no longer completed any competence tests. In a second approach, about 380 children took part in the assessment by filling out only questionnaires distributed by their class teachers within the class context. For the third approach, the 920 children who had been in the study since elementary school were supplemented by an additional 879 new classmates (see Schmidt, Schmitt, & Smidt, 2009; Kurz, Kratzmann, & von Maurice, 2007). This sample was given questionnaires as well as competence tests to fill out, and therefore serves as the basis of the following sample description. Most students in this sample (62%) attended the Gymnasium then, 18% went to Realschule, and 21% chose the Hauptschule. Altogether, BiKS-8-14 had a total sample size of nearly 3,000 students.

Figure 3 shows the panel attrition across the seven measurement points from the year 2006 on. Similar to the first sample described above, only the sample size based on parental permission is shown regardless of the number of students who were absent on the test day. Generally, there was a decline across time as usually found in empirical research. The decrease in 2011 was – as happened in the other sample – due to the parents who declined to renew their permission. In this case, not only the parents had to agree to the further participation of their child, but the students themselves were also asked for their permission because most of them had reached the age of 14. At this age, the students had to be asked personally according to German law. Not surprisingly, a substantial part of the sample refused to give their permission. The remaining sample consisted of almost 2,000 students.

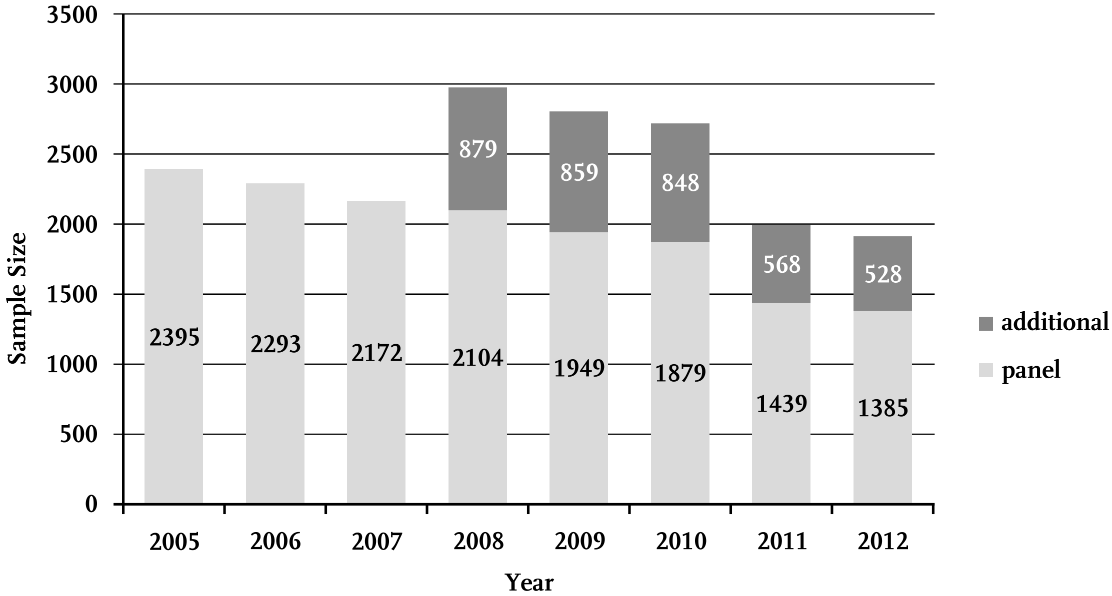

Figure 3. Sample enhancement and panel attrition for BiKS-8-14.

Measurement methods

A variety of different information was surveyed in the BiKS study. At each of the measurement points (i.e., 14 in BiKS-3-10 and nine in BiKS-8-14; cf. Figure 1), multiple instruments were applied. These instruments can be roughly divided into instruments related to or applied in institutional settings (e.g., questionnaires for preschool teachers and teachers, monitoring instruments, and competence tests administered in individual or group settings for the children and students), instruments related to or applied in family settings (e.g., questionnaires and computer-assisted telephone interviews for parents, monitoring instruments in the family context, and competence tests as individual tests for the children and students); in addition, qualitative interviews (personal interviews with teachers, parents, or children) were conducted. Some of the instruments were applied only to a subsample (e.g., only to Bavarians or only to some Turkish participants). Due to the frequent observations and repetition of the same or similar instruments, the BiKS data provide an excellent opportunity to trace the children's development very closely and to explore the conditions of this process with a unique variety of factors.

To provide deeper insight into the variety of competence instruments applied in the BiKS study, the two following tables present the main competence facets that were surveyed over time. Whereas BiKS-3-10 focused initially on language acquisition and cognitive development in kindergarten as precursors of the academic competencies measured from the first grade on, BiKS-8-14 naturally had academic measures at the center of its research from the beginning. The competence tests that were used consisted of either self-developed and piloted or established instruments. All tests were chosen to be appropriate for the children's age at each measurement point and allow for comparability over time.

In the following, Table 3 provides an overview of the competence facets that were assessed at each of the main measurement points of BiKS-3-10. They were given as either individual tests in the family context or as group tests in schools. Some of the measures (e.g., reading comprehension) were the same as in the second sample, BiKS-8-14, so that the children of the two samples could be linked to the time when each of the cohorts attended the fourth grade of elementary school. Competence facets were not necessarily measured with one and the same competence test, even if they are named equal across the measurement points in the table.

Table 3. Main Competence Facets Measured in the BiKS-3-10 Sample across Time*

	2005 1st year kinder-garten	2006 2nd year kinder-garten	2007 3rd year kinder-garten	2008/09 1st grade elementary school	2009/10 2nd grade elementary school	2010/11 3rd grade elementary school	2011/12 4th grade elementary school
Language	vocabulary	vocabulary	vocabulary	vocabulary	vocabulary	vocabulary	vocabulary
	grammar	grammar	grammar	grammar	grammar	grammar	
	indicators of language production	indicators of language production	indicators of language production				
				academic language indicators	academic language indicators	academic language indicators	academic language indicators
Reading				reading compre-hension	reading compre-hension	reading compre-hension	reading compre-hension
					reading speed	reading speed	reading speed
Working memory	verbal short term memory	verbal short term memory	verbal short term memory			verbal short term memory	verbal short term memory
	nonverbal short term memory	nonverbal short term memory	nonverbal short term memory				
Knowledge	factual and conceptual knowledge	factual knowledge	factual knowledge				
Speed of information processing		naming speed	naming speed	naming speed	naming speed		
Nonverbal cognitive abilities	abstract reasoning	abstract reasoning	abstract reasoning	abstract reasoning	abstract reasoning	abstract reasoning	abstract reasoning
Mathe-matics	arithmetic	arithmetic	arithmetic	arithmetic	arithmetic	arithmetic	arithmetic
Indicators of meta-cognitive under-standing		meta-cognitive under-standing	meta-cognitive under-standing	meta-cognitive under-standing	meta-cognitive under-standing	meta-cognitive under-standing	meta-cognitive under-standing

*The table includes only the main assessment points (when all children were tested) and only central measures; some measurement instruments had to be changed according to age and measurement point. Some of the competence facets were, at some measurement points, assessed by various instruments/indicators and some were only gathered from subsamples.

Similar to the previous table above, Table 4 displays the competence measures of BiKS-8-14, starting in the third grade of elementary school in 2006 and going to the end of secondary school. The focus of this sample was on academic achievement and reading competence, Thereby, the development of these competencies can be described across an 8-year period, including various factors collected by the other instruments that were applied.

Table 4. Main Competence Facets Measured in the BiKS-8-14 Sample across Time*

	2005/06 3rd grade elementary school	2006/07 4th grade elementary school, 1st term	2006/07 4th grade elementary school, 2nd term	2007/08 5th grade secondary school	2008/09 6th grade secondary school	2009/10 7th grade secondary school	2010/11 8th grade secondary school	2011/12 9th grade secondary school
Language	vocabulary	vocabulary	vocabulary	vocabulary	vocabulary	vocabulary	vocabulary	vocabulary
		grammar						
	listening compre-hension							
				foreign language English	foreign language English	foreign language English		
Reading	reading compre-hension	reading compre-hension	reading compre-hension	reading compre-hension	reading compre-hension	reading compre-hension	reading compre-hension	reading compre-hension
	reading speed							
Writing	ortho-graphy		ortho-graphy		ortho-graphy	ortho-graphy		
Non-verbal cognitive abilities	abstract reasoning	abstract reasoning	abstract reasoning	abstract reasoning	abstract reasoning	abstract reasoning	abstract reasoning	abstract reasoning
Mathe-matics	arithmetic	arithmetic	arithmetic	arithmetic, geometry+ story problems	arithmetic, geometry+ story problems	arithmetic, geometry+ story problems	arithmetic, geometry+ story problems	

*Some of the competence facets were, to some measurement points, only gathered from subsamples.

Furthermore, both the BiKS-3-10 and BiKS-8-14 studies were specifically amended by several qualitative and quantitative surveys with different subsamples that are not depicted here separately. Such a detailed examination that follows children from age 3 to age 15 is unique in the field of educational research and, as this book demonstrates, provides a wide variety of options for analyses.

References

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

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London & New York: Routledge.

Kurz, K., Kratzmann, J., & von Maurice, J. (2007). *The BiKS study. Method report on the sampling procedure*. Online available via: http://psydok.sulb.uni-saarland.de/volltexte/2007/990/index.html (called on 18.06.2012).

OECD (2010). PISA 2009 Results: Executive Summary.

Schmidt, S., Schmitt, M., & Smidt, W. (2009). *The BiKS study. Method report on the second phase*. Online available via: http://psydok.sulb.uni-saarland.de/volltexte/2009/2534/ (called on 18.06.2012).

Chapter 3

3 Facets of Preschoolers' Home Literacy Environments: What Contributes to Reading Literacy in Primary School?

Simone Lehrl, Susanne Ebert, and Hans-Guenther Rossbach

Summary

How the family makes early contributions to the acquisition of children's emergent literacy skills and later reading literacy has received increased attention throughout the research literature. Numerous studies have accumulated evidence for the relation between the home literacy environment (HLE) when children are of preschool age (e.g., shared book reading interactions) and children's literacy and language skills. In order to understand how the HLE shapes children's reading literacy before formal schooling actually begins, it is important to examine how specific aspects of the HLE contribute to the development of children's reading literacy. After a short review of the existing research regarding the influence of the HLE on children's reading literacy, the current chapter presents findings from the BiKS-3-10 study. Many studies focus on only one specific aspect of the HLE – mainly, the frequency of shared book reading – at only one time point across the

Author Note

Simone Lehrl,
Early Childhood Education, University of Bamberg, Germany.

Susanne Ebert,
Department of Developmental Psychology, University of Bamberg, Germany.

Hans-Guenther Rossbach,
Early Childhood Education, University of Bamberg, Germany.

This research was supported by grant RO 820/11-1, 11-2 & 11-3 from the German Research Foundation (DFG).

Correspondence concerning this chapter should be addressed to Simone Lehrl, Early Childhood Education, University of Bamberg, Luitpoldstrasse 5, 96052 Bamberg, Germany. E-mail: simone.lehrl@uni-bamberg.de

preschool period. By contrast, the present study examined the relation between various specific home literacy practices for children of preschool age (book exposure, formal instruction, and the quality of parent-child interactions during shared book reading) and children's reading literacy (basic reading skills and reading comprehension) in elementary school by using different measures (questionnaires and observations) at different measurement points. Results showed that the different aspects of the HLE were significantly related to the different aspects of reading literacy in elementary school. Furthermore, mediation analyses revealed that the effects of the HLE on reading literacy were mediated through emergent literacy skills. The findings underline the importance of the home literacy environment and indicate that research approaches should be strengthened by using multiple measures of the home literacy environment.

Theoretical Background

Reading is known to be one of the most essential competencies that are needed for people to successfully participate in society (OECD, 2003). Although reading is supposed to be acquired via formal instruction in school, we know that children have a lot of experiences with written language before formal schooling begins. Children are surrounded by letters and words in everyday life, beginning with their written name on the front door. They see adults reading newspapers and books and begin to understand that there may be meaning behind the written signs. These kinds of early experiences with written language begin to form the knowledge and skills that are crucial for later reading development. These precursors of later reading are known as emergent literacy (Whitehurst & Lonigan, 1998) and include knowledge of the reading process and of letters, of phonological information processing such as phonological awareness, as well as oral language and linguistic abilities including vocabulary and grammar (see also Ebert & Weinert, chapter 5, this volume). However, what we know so far is that there is a great deal of variability in this knowledge and in these skills and that these individual differences are related to the social backgrounds of the families (Dubowy, Ebert, von Maurice, & Weinert, 2008; Weinert, Ebert, & Dubowy, 2010; Weinert & Ebert, in press).

Presumably one source of variance is comprised of a child's experiences at home and at preschool. For educational research, it is therefore essential to understand the factors in the home learning environment that influence emergent literacy and later reading literacy. Various studies have demonstrated that the home learning environment is associated with children's early literacy and reading development (e.g., Melhuish, et al., 2008; Son & Morrison, 2010; Ebert, et al., 2012; Weinert, Ebert, Lockl, & Kuger, 2012). The most considered variable in this context is the frequency with which parents read to a child (Burgess, Hecht, & Lonigan, 2002). Although a meta-analysis by Bus, van Ijzendoorn, and Pellegrini (1995) showed positive effects of the frequency of reading to a child on emergent literacy (e.g., letter knowledge) as well as on oral language skills (e.g., vocabulary), the amount of explained variance was only moderate (see also Scarborough & Dobrich, 1994). These moderate effects caused some researchers to challenge whether the frequency of reading to a child was suitable or sufficient for explaining interindividual differences in the ability to acquire reading literacy; thus, they suggested extending the concept of the home learning environment (e.g., Burgess, et al., 2002; van Steensel, 2006). In this vein, the following chapter investigates the meaning of different facets of the early home learning environment for later reading literacy.

The Family's Contribution to Reading

The family is the first environment the child encounters and therefore seems to be an important source for children's development. Accordingly, with regard to reading literacy, the early home learning environment – also known as the home literacy environment (HLE) in the research on literacy development – is known to affect the competencies that are necessary for an individual to learn to read in a conventional way; these competencies are called *emergent literacy* (Whitehurst & Lonigan, 1998). Emergent literacy is a term used to describe young children's development with regard to written language (Whitehurst & Lonigan, 1998). An essential aspect of this definition is that this process is continuous and begins long before formal instruction in school begins (Teale & Sulzby, 1989). The skills included in the emergent literacy concept are oral language skills, phonological awareness, knowledge of letters, and perceptions of print (Whitehurst & Lonigan, 1998). All these skills have been shown to be of great importance for later reading development (e.g., Ebert & Weinert, chapter 5, this volume). Accordingly, the HLE comprises the resources and opportunities the family offers to the child regarding written and oral language (Burgess, et al., 2002). However, there is no well-accepted definition or operationalization of the HLE. This has led to a wide variety of operationalizations of the concept ranging from single-item approaches to as many as 10 different dimensions (Leseman & de Jong, 1998; Britto & Brooks-Gunn, 2001; Umek, Podlesek, & Fekonja, 2005; Gonzalez, et al., 2011;). Sénéchal, LeFevre, Thomas, and Daley (1998) suggested a theory-driven approach that distinguishes between informal and formal literacy activities at home – called the home literacy model. Whereas formal literacy activities at home refer directly to print and are reflected by, for example, teaching the sounds corresponding to certain letters, informal literacy activities refer to experiences that are not focused on print per se but rather on the contents of printed material. These informal experiences are gained specifically through story book exposure. Story book exposure is usually measured by the number of books owned and the amount of time spent reading with or to a child. The authors showed that the two dimensions are distinct from each other as they were not correlated and varied in their prediction of emergent literacy skills. The home literacy model provides specific assumptions concerning the relation that each dimension has to the development of reading literacy.

Formal literacy experiences at home

Formal literacy experiences are assumed to foster reading skills, such as word decoding, which occurs through the fostering of early letter knowledge and early word reading skills (Sénéchal & LeFevre, 2002). Formal literacy experiences are most often measured by having a child state the alphabet, write his or her own name, and read simple words. Such formal experiences have been shown to be associated with letter knowledge (Evans, Shaw, & Bell, 2000; Torppa, Poikkeus, Laakos, Eklund, & Lyytinen, 2006; Manolitsis, Georgiou, Stephenson, & Parrila, 2009; Lehrl, Ebert, Rossbach, & Weinert, 2012) and word decoding skills (Sénéchal & LeFevre, 2002; Hood, Conlon, & Andrews, 2008). Lehrl and colleagues (2012), for example, found that the (self-reported) frequency with which parents taught their child to read and to recite the alphabet at the age of 3 years predicted letter knowledge at the age of 6, even when earlier language competencies were controlled. Similarly, Torppa and colleagues (2006) found that the frequency with which parents taught letter names when their child was 4.5 years old predicted the child's letter knowledge at the age of 6. Other studies have even shown that such formal activities also have substantial effects on later, more advanced reading skills (see Scarborough & Dobrich, 1994, for a review; Evans, et al., 2000; Sénéchal & LeFevre, 2002; Whitehurst & Lonigan, 2003; Sénéchal, 2006; Stephenson, Parrila, Georgiou, & Kirby, 2008). For example, in an English-speaking sample, Sénéchal and LeFevre (2002) demonstrated that parental reports of how often they taught reading and writing to their kindergarten-aged children were indirectly linked to word reading skills in Grade 1 through emergent literacy skills. In the same manner, parental reports of how often they taught literacy skills were also related to word reading skills in Grade 3. The same was true for a French-speaking sample (Sénéchal, 2006). However, no such effects were found in a Greek sample by Manolitsis et al. (2011) or in a Finnish sample by Leppaenen, Niemi, Aunola, & Nurmi (2004). These findings suggest different effects for different languages, depending on their orthographical transparency (i.e., the extent to which graphemes have multiple pronunciations and phonemes have multiple spellings; Manolitsis, et al., 2009). Presumably the teaching of the sounds of letters before formal schooling begins is especially important for children who are learning written languages that are orthographically less transparent (e.g., French, English) and when reading acquisition is more difficult (Georgiou, et al., 2008). As German is an orthographically transparent language, we assume that the effects of formal teaching might be low or even absent with regard to reading literacy.

Informal literacy experience at home

According to the home literacy model, informal literacy experiences are assumed to promote language skills, especially vocabulary, and in accordance, these language skills then promote early reading literacy. As Sénéchal's (2006) study focused on more advanced reading skills, her findings suggest indirect effects of informal literacy experiences via vocabulary on reading comprehension. Book exposure and shared book reading in particular can be seen as the prototypical aspect of informal literacy experience. In the context of shared book reading, children are exposed to oral language, print, and literacy concepts (Sénéchal & LeFevre, 2001). The importance of shared book reading has been investigated a lot and has consistently shown positive effects on language and literacy skills (i.e., children's vocabulary development, phonemic skills, print concept knowledge, and positive attitudes toward literacy; Dickinson & Tabors, 1991; DeBaryshe, 1993; Wagner, Torgesen & Rashotte, 1994; Sénéchal, LeFevre, Hudson, & Lawson, 1996; Lyytinen, Laasko, & Poikkeus, 1998; Raikes, et al., 2006). The meta-analysis by Bus et al. (1995) indicated that the amount of shared book reading was related to children's language skills, emergent literacy skills, and reading skills (see also Scarborough & Dobrich 1994). Additionally, some results have indicated that the number of picture books in a home is positively associated with children's language and reading skills (e.g., Payne, Whitehurst, & Angell, 1994; Sénéchal, et al., 1996; Sénéchal, et al., 1998). As these aspects cover the frequency of shared book reading and number of books, this dimension could be titled *quantity* of book exposure.

When thinking about how book exposure contributes to children's literacy development, a social-constructionist perspective suggests that books are a source from which children can acquire literacy skills while being supported by a more experienced person (Wygotsky, 1969). Consequently, it is assumed that children become interested in books, expand their vocabulary, and acquire other emergent literacy skills through the social interaction that occurs during the shared reading experience. Young children may profit from the guidance of an experienced reader with regard to understanding the meaning behind the print (Snow, Burns, & Griffin, 1998). Thereby, book reading seems to be most effective when parents actively involve their child in the reading situation by asking open-ended questions (Ninio, 1983), discussing the story, and elaborating on the child's comments in verbal exchanges (de Jong & Leseman, 2001). This assumption is also supported by research that has investigated the effects of reading interventions. The benefit of a reading intervention that emphasizes the interactive style of reading on young children's language skills was demonstrated first by Whitehurst and his colleagues (1988). The so-called Dialogic Reading Program was designed to encourage the parents of 2- and 3-year-old children to use evocative techniques that encourage the child's active participation in telling the story by asking questions and by using expansions, corrections, and praise to give the child feedback

(Arnold & Whitehurst, 1994). Furthermore, a meta-analysis by the Early Literacy Council showed higher effect sizes for child reading outcomes in interventions that were designed to include the child in an interactive way in the reading situation than in interventions with less emphasis on the interactive involvement of the child (Shanahan & Lonigan, 2010). These findings support the idea that, in addition to examining the *quantity* of book exposure, researchers should also examine parent-child interactions while book reading. Thus, the *quality* of book exposure can be seen as a second informal dimension of the HLE. Lehrl and colleagues (2012) showed that the quality of parent-child interactions in a shared book reading situation (e.g., asking open-ended questions and using complex language) measured when the children were about 3 years of age, explained unique variance in the growth of the children's vocabulary in the next year, whereas the quantity of book exposure explained unique variance in the growth of grammatical knowledge in the same time period. Similar findings regarding the differential effects of the quantity and quality of book exposure can be found in a Dutch study conducted by Leseman and de Jong (1998). They reported that the quality of instruction while sharing a book with a preschooler was positively associated with vocabulary development at the age of 7, whereas the aspect that reflected quantity – literacy opportunity – was not.

In summary, a distinction between formal and informal dimensions of the HLE as assumed by the home literacy model is consistent with research findings from different samples. However, an extension of a further informal dimension that refers to the quality of parent-child interactions seems to be necessary. In light of this and to provide an extension to Sénéchal and LeFevre (2002), the current study accordingly distinguished between three dimensions of the HLE: formal instruction in literacy, book exposure (quantity), and the quality of parent-child interactions. Lehrl and colleagues (2012) showed that each dimension was associated with different emergent literacy outcomes at the age of 4 years. The present study extended these findings by focusing on the same children at an older age and by employing reading literacy outcome measures. The main question was whether the three facets of the HLE would also have differential effects on reading literacy. Thereby, our study differentiated between different aspects of reading literacy. This approach is theoretically driven by Snow's (1991, 1999) componential model of literacy development in school. This model suggests that basic reading skills such as word decoding or reading speed and reading comprehension comprise two different but interrelated facets of reading literacy that are determined by different environmental and cognitive preconditions (see also Scarborough, 2001; Richter & Christmann, 2002; Storch & Whitehurst, 2002). Of course, the two dimensions are interrelated as at least a minimum of basic reading skills are necessary for reading comprehension (Hoover & Gough, 1990). According to the model, basic reading skills are determined in particular by code-related emergent literacy skills such as letter knowledge and phoneme awareness. These in turn are assumed to be fostered by home literacy experiences that are focused directly on teaching the alphabet and print-related skills. By contrast, reading comprehension will be specifically affected by children's vocabulary, world knowledge, and pragmatic skills, which in turn are assumed to be predicted by informal literacy experiences such as story book exposure. Against this background, the current study addressed the following questions:

1) Does each aspect of the HLE explain unique variance in children's reading literacy beyond the others?
2) Do the various aspects of the HLE have a different impact on reading comprehension in comparison to basic reading skills?
3) Are the effects of the HLE mediated by emergent literacy skills at the end of preschool?

Method

Procedure and Sample

All data for the present study were drawn from the BiKS-3-10 substudy (see also Lorenz, Schmitt, Lehrl, Mudiappa, & Rossbach, chapter 1, this volume). At the first measurement point in autumn 2005, a sample of 547 children (about 3 years old) attending 97 preschools in two German federal states (Hesse and Bavaria) participated. Data collection took place in half- or 1-year intervals and contained a wide range of data on child and family characteristics as well as data on their learning environments at home, in the preschools, and in the primary schools.

The present study focuses on children's reading literacy in the second grade of primary school. Because not all children could be followed over such a long period of time, the sample size was reduced to 343 children for whom at least one outcome measure in reading literacy in Grade 2 was available. The average age of the children was 8.2 years (SD = 0.33) in Grade 2. Furthermore, the gender of the children was nearly equally distributed; 48.4% were male and 51.6% were female.

Measures

Reading literacy. Reading literacy was assessed using a test that measures basic reading skills, specifically reading speed, as well as a test of reading comprehension. Both tests were administered in the second grade of primary school when the children were about 7 years old.

Basic reading skills. The SLS 1-4 (Salzburger Lese-Screening fuer die Klassenstufen 1-4; Mayringer & Wimmer, 2003; parallel test reliability > .90) assesses reading speed as a measure of basic reading skills. It consists of a list of 70 short simple statements (e.g., "Bananas are blue"), and children have to read as many sentences as possible in 3 min. Thereby, children have to mark whether the statements, ordered by increasing length, are true or false. The dependent variable is the sum of the correctly classified sentences (M = 32, SD = 10).

Reading comprehension. To assess reading comprehension, the subtest "text comprehension" of the ELFE 1-6 (Ein Leseverstaendnistest fuer Erst- bis Sechstklaessler; Lenhard & Schneider, 2005; retest reliability r > .90) was administered.

For this subtest, students have to read 20 short passages on various topics, mainly of everyday life, and then have to answer comprehension questions in a multiple-choice format. The dependent variable is the sum of the correct responses (M = 10, SD = 4).

Emergent literacy. All emergent literacy competencies were measured in the final year of preschool when children were about 5 years of age. For this study, we focused on children's receptive vocabulary, receptive grammar, and letter knowledge.

Receptive vocabulary. To assess children's receptive vocabulary, a German research version of the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1981) was used. For each item, the child was required to choose the picture that represented a verbally given word out of four alternatives. The test had 175 items of increasing difficulty. Testing was stopped when six or more items within a set of 12 items were answered incorrectly. The indicator for receptive vocabulary consisted of the sum score of all correct items ($M = 80$, $SD = 21$).

Receptive grammar. A shortened German Version of the Test for the Reception of Grammar (TROG; Bishop, 1989; German Version TROG-D, Fox, 2006) was used. The test consists of 48 items ordered in sets of four or two items and requires the child to select the picture that corresponds to a given sentence (out of four alternatives). Testing was stopped when children answered five succeeding sets incorrectly (a set was counted as incorrect when at least one item of a set was answered incorrectly). The sum score of all correct items was used to build an indicator for receptive grammar ($M = 37$, $SD = 5$).

Letter knowledge. Children were exposed to the 26 letters of the German alphabet in five or six letter groupings depicted on cards (20 x 15 cm). Letters had a height of 2 cm and were grouped together incidentally. We ensured that no letter was followed by the letter that immediately followed it in the alphabet. On each picture card, the children were asked to name the letters they knew. The formal as well as the phonemically correct pronunciation were scored as correct answers. The sum of all correctly named letters was used in the analyses ($M = 13$, $SD = 8$).

The Home Literacy Environment and family background

The Home Literacy Environment (HLE). The HLE was measured in the first, second, and third year of preschool education. According to our research question, we differentiated between three facets of the HLE (formal instruction, book exposure, and the quality of parent-child interactions). Each measure was calculated by taking the mean of the three yearly measurement occasions.

Formal instruction. Parents were asked to report the frequency with which they taught their child to read and to recite the alphabet on a 4 point scale (1 = *never*, and 4 = *very often*). Both items were taken out of the Home Observation for Measurement of the Environment (HOME; Caldwell & Bradley 1984): “The child is encouraged to learn to read a few words.”, “The child is encouraged to learn the alphabet.”. The correlation between the two items at each measurement occasion was *r* = .77, *r* = .72, and *r* = .71, respectively.

Book exposure. Book exposure was measured via the answers the parents gave in a questionnaire regarding how frequently they read to the child (1 = *never*, and 5 = *daily*), the number of books in the household, and the number of children's books in the household. Regarding the books in the household, categories ranged from 1 = *up to 30*, 2 = *up to 100*, 3 = *up to 200*, and 4 = *more than 200 books*. The categories for children's books ranged from 1 = *up to 10*, 2 = *up to 20*, 3 = *up to 30*, and 4 = *more than 30 books*. In order to represent one scale, before taking the means of the items, the items were first standardized. Cronbach's alpha for each measurement occasion was .68, .67, and .70, respectively.

Quality of parent-child interactions. The Family Rating Scale (Familieneinschaetzskala (FES); Kuger, Pflieger, & Rossbach 2005), developed in the context of the BiKS study, was used to measure the quality of parent-child interactions during a semi-standardized book reading task between the primary caregiver (96% were mothers) and the child. The book provided by the research team was not commercial and therefore unknown to all of the parents. The interaction between parent and child was rated on 11 general and domain-specific aspects of interaction quality (1 = *low quality* to 7 = *high quality*) by trained observers. As a measure of the quality of parent-child interactions in the present study, the following items were used: use of questions when interacting, quality of oral language, verbal distancing, nonverbal behavior, participation in dialogue, and use of phonological cues. Cronbach's alpha for each measurement occasion was .65, .75, and .77, respectively.

Native language background. Parents were asked what their first language was. In 17.2% of the families in the present subsample, at least one parent indicated a mother tongue other than German.

Socioeconomic status of the family (SES). SES was measured using the International Socioeconomic Index of Occupational Status (ISEI; Ganzeboom & Treiman, 1996). The highest value (HISEI) of each family was used in the analyses (range: 16 - 90; *M* = 53.1; *SD* = 16.1).

Statistical Analyses

To examine the impact of the home learning environment on children's reading literacy, path models were run. To answer the first two research questions regarding the impact of the various measures of the HLE on reading literacy, a test of a path model involving the two outcome measures (basic reading skills and reading comprehension) was conducted (see Figure 1).

According to the theoretical background, we expected effects of the HLE on emergent literacy skills, which were then, according to the home literacy model, expected to predict reading literacy. In order to answer our third research question regarding whether the effects of the HLE would be mediated through emergent literacy skills, an additional path model was specified including the variables vocabulary, grammar, and letter knowledge as indicators of emergent literacy in the final preschool year. This made it possible to test for indirect effects of the HLE on reading literacy through emergent literacy competencies. A full mediation model (without direct paths) as well as a partial mediation model (allowing direct paths) was tested. The chi-square difference test was used to find the best-fitting solution, which is displayed in Figure 2.

Mplus version 6.0 (Muthén & Muthén, 1998-2010) was used for all analyses. Model fit was evaluated by the chi-square test, RMSEA, SRMR, and CFI, as recommended by Hu and Bentler (1999). The amount of missing data for the single predictors of interest in the sample was very small (9.4% on average; ranging from 0% to 27.7%). In an attempt to avoid introducing bias into the sample through listwise deletion (Little & Rubin, 1987), the full-information maximum likelihood (FIML) approach (Arbuckle, 1996), which includes valid information for all observations for model estimation, was used to deal with missing data.

Results

Relations between the HLE and Reading Literacy

Table 1 displays the bivariate correlations between the HLE measures and reading literacy. The results indicated significant relations between the HLE and reading literacy. The correlations supported the proposed pattern that formal instruction would be associated with basic reading skills (r = .16), whereas reading comprehension was more strongly correlated with the informal dimensions: the quality of parent-child interactions ($r = .20$, $p < .05$) and book exposure ($r = .21$, $p < .05$).

Table 1. Correlations between Background Variables, HLE, Emergent Literacy, and Reading Literacy

	1.	2.	3.	4.	5.	6.	7.	8.	9
1. Reading comprehension									
2. Basic reading skills	.78**	-							
3. HLE interaction quality	.20**	.02	-						
4. HLE book exposure	.21**	.24**	.26**	-					
5. HLE formal instruction	.07	.16*	.09	.00	-				
6. Vocabulary	.36**	.16*	.30**	.27**	-.04	-			
7. Grammar	.36**	.21**	.32**	.31**	-.08	.62**	-		
8. Letter knowledge	.41**	.39**	.01	.06	.29**	.18*	.10	-	
9. SES	.17*	.22**	.33**	.41**	-.11	.39**	.36**	.16**	-
10. Native language background	-.12#	-.10#	-.30**	-.16*	.10	-.46**	-.33**	.05	-.19*

Note. Language background: 0 = both parents German, 1 = one parent not German.
SES = socio-economic status.
# $p < .10$. * $p < .05$. ** $p < .01$.

As mentioned earlier, theory suggests that basic reading skills are a necessary prerequisite for reading comprehension. Thus, in the path model, the variable "basic reading skills" was regressed on the variable reading comprehension. Furthermore, all background and HLE variables were regressed on the outcome measures. The predictor variables were allowed to correlate. The resulting path model (see Figure 1) tested whether and what impact literacy experiences at home in the preschool years have on reading literacy in the second grade of primary school, when considered simultaneously. It demonstrated that, while controlling for background variables, story book exposure was significantly associated with reading comprehension, even after controlling for basic reading skills (ß = .09, $p < .10$).

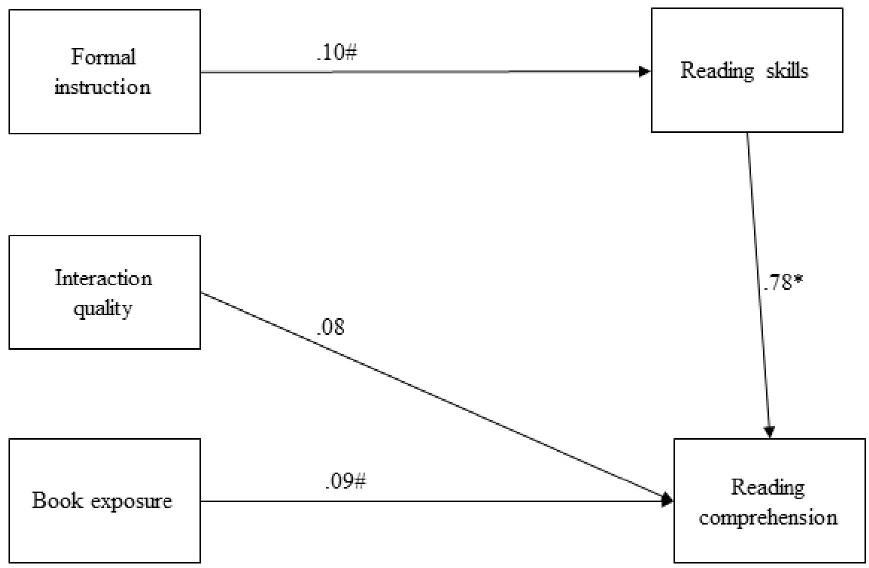

Figure 1. The relation between facets of the home literacy environment and reading literacy. *Note.* $N = 343$, $\chi^2(df) = 2.57(3)$, $p = .46$, CFI = 1.00, RMSEA = .00, $p = .75$, SRMR = .01. SES and native language background were controlled in the path model.
# $p < .10$. * $p < .05$.

Against our expectations, the other informal facet of the HLE, the quality of parent-child interactions, did not explain significant unique variance in reading comprehension (ß = .08, *ns*). However, as expected, parents' formal instruction had no significant effect on reading comprehension, but had a marginally significant effect on basic reading skills (ß = .10, $p < .10$). Furthermore, neither story book exposure nor the quality of interactions predicted basic reading skills. We also found that there was a strong association between basic reading skills and reading comprehension (ß = .78, $p < .01$). The explained variance for reading comprehension was correspondingly high ($R^2 = .61$) and comparatively low for basic reading skills ($R^2 = .07$). Thus, the direct effects of the HLE on reading literacy were relatively small.

Indirect Effects of the HLE via Emergent Literacy

In a second step, a path model that predicted emergent literacy skills was specified to ascertain whether early language competencies would mediate the effects of the HLE on reading literacy. Concerning the effects from the HLE on emergent literacy, we specified the paths according to the theoretical assumptions. Thus, a path leading from formal instruction to letter knowledge was specified. Furthermore, paths leading from book exposure to vocabulary and grammar were specified as well as paths leading from the quality of interactions to vocabulary and grammar. Additionally, paths were specified leading from preschool skills to both measures of reading literacy from Grade 2 (i.e., reading comprehension and basic reading skills).

As our focus was on the direct and indirect links between the HLE measures and reading literacy, a first model that allowed only indirect effects (full mediation) was compared to a second model that also allowed direct effects (partial mediation). In the partial mediation model, none of the direct effects were significant. Accordingly, the full mediation model did not show a worse fit than the partial mediation model as the chi-square difference test demonstrated ($\Delta\chi^2$ = 6.8, *df* = 3, *p* = .08). Thus, the full mediation model as the more parsimonious was preferred. Figure 2 shows all significant indirect effects in this model with bolt arrows.

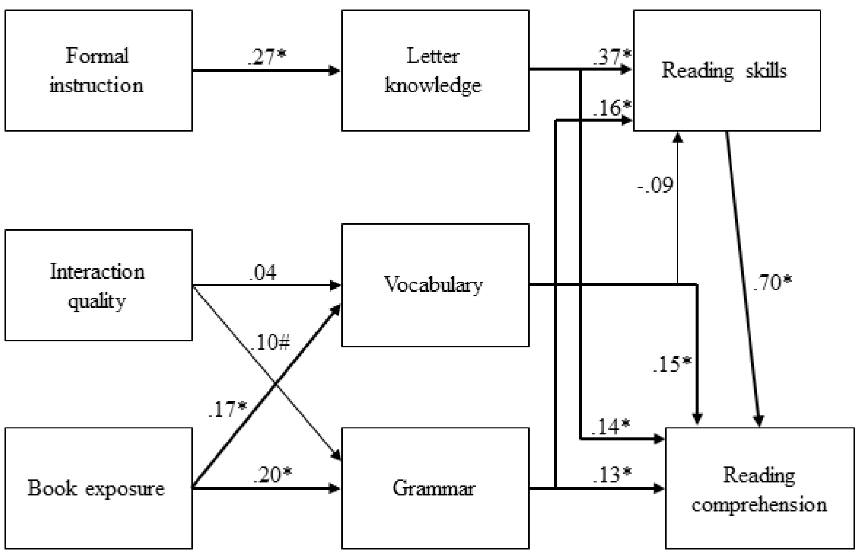

Figure 2. The relation between facets of the home literacy environment and reading literacy including mediating variables (i.e., vocabulary, grammar, and letter knowledge). *Note.* $N = 343$; $\chi^2(df) = 8.53(10)$, $p = .57$, CFI = 1.00, RMSEA = .00, $p = .94$, SRMR = .02. SES and native language background were controlled in the path model.
# $p < .10$. * $p < .05$.

Concerning basic reading skills, the effect of formal instruction was completely mediated through letter knowledge (indirect effect: ß = .10, *p* < .05). Additionally, the indirect effect of book exposure through grammar (ß = .03, *p* < .10) on basic reading skills was significant. However, there was no significant indirect effect on basic reading skills through vocabulary (ß = -.02, *ns*). Furthermore, no indirect effects of the quality of interactions were found on basic reading skills through vocabulary (ß = -.00, *ns*) or through grammar (ß = -.02, *ns*).

The indirect paths representing the effects of book exposure on reading comprehension – controlling for basic reading skills – through grammar (ß = .03, $p < .05$) and vocabulary (ß = .03, $p < .05$) were significant. An additional indirect effect of book exposure was found through grammar and basic reading skills (ß = .02, $p < .10$). The other hypothesized indirect path leading from book exposure through vocabulary and basic reading skills to reading comprehension was not significant (ß = -.01, *ns*). The same was true for interaction quality where no indirect path approached significance. Surprisingly, the formal instruction of the parents showed a significant indirect effect via letter knowledge (ß = .04, $p < .05$), although there was no association between formal instruction and reading comprehension in the model without the mediating variable letter knowledge (see Zhao, Lynch & Chen, 2010, for a discussion on testing mediation when no zero-order correlation exists). The variable letter knowledge therefore seemed to act as a suppressor variable that cloaked the relation between reading comprehension and formal instruction. Furthermore, the indirect path leading from formal instruction through letter knowledge and basic reading skills was significant as well (ß = .07, $p < .05$) and even higher than the effect that went through only letter knowledge.

Discussion

In the present longitudinal study, the complex relations between preschoolers' home literacy environments, developing literacy skills, and reading literacy in Grade 2 were examined. Three measures of the early home learning environment representing formal and informal stimulation at home – formal instruction, book exposure, and the quality of parent-child interactions as well as different measures of reading literacy in Grade 2 (i.e., basic reading skills and reading comprehension) were investigated. Furthermore, selected emergent literacy competencies (i.e., grammar, vocabulary, and letter knowledge) were taken into account. The main results of the study are: There are relations between the early home learning environment and reading literacy in

Grade 2. These effects are (a) different when considering the different dimensions of the home learning environment, (b) different for basic reading skills and reading comprehension, and (c) mediated through emergent literacy skills in preschool.

The first aim of the study was to examine the contributions of different aspects of the HLE on basic reading skills and reading comprehension. Our findings are in line with previous studies and extend them by showing that formal instruction by the parents such as the explicit teaching of reading-related skills was associated not just with decoding skills (Sénéchal & LeFevre, 2002; Sénéchal, 2006; Manolitsis, et al., 2009) but also with basic reading skills. Experiences that included informal interactions with print on a more global level such as reading to the child and having a lot of books appropriate for both children and adults were associated with reading comprehension (Leseman & de Jong, 1998; de Jong & Leseman 2001; Sénéchal, 2006). Also consistent with previous findings, most of the effects could be attributed to the effects of the HLE on emergent literacy competencies (i.e., letter knowledge, vocabulary, and grammar in the final year of preschool). This had an effect on reading literacy in particular when parents indicated that they frequently taught the alphabet and frequently had their children read simple words. This effect was mediated through letter knowledge for both basic reading skills and reading comprehension. Nevertheless, the effect of formal instruction via letter knowledge was stronger for basic reading skills than for reading comprehension. From a theoretical point of view, the results correspond to the home literacy model as well as Snow's (1991, 1999) two-trajectory model of literacy development in school. The specific relation found in the present study between formal instruction in the home and basic reading skills in the second grade via letter knowledge in preschool is in line with the model as it suggests that code-related skills will be specifically affected by home literacy experiences that refer to print. The present study demonstrates that basic reading skills that are more focused on reading speed and less on decoding are also affected by home literacy experiences that refer to print. The importance of letter knowledge for reading comprehension may reflect the idea that even in the second year of formal reading instruction, reading comprehension demands a lot of basic reading skills (Perfetti, 1985; see also Ebert & Weinert, chapter 5, this volume). This concept is demonstrated by the strong correlations between basic reading skills and reading comprehension. However, the present study modelled basic reading skills and reading comprehension simultaneously under the assumption that reading comprehension is directly influenced by basic reading skills. Nevertheless, the indirect effect of formal instruction on reading comprehension via letter knowledge remains, even when controlling for basic reading skills. Thus, formal instruction by an experienced other has an effect through letter knowledge on reading comprehension over and above basic reading skills. Reasons for why letter knowledge is such a crucial skill in reading development has been summarized by Foulin (2005), who states: “[...] LNK [letter-name knowledge] may set prereaders on the right path towards conventional alphabetical literacy” (p. 136). To summarize, formal instruction with regard to letters by the parents could help the child to get to know the letters earlier and seems to boost their reading literacy. As Sénéchal and LeFevre (2002) pointed out, the exposure to books may not be sufficient to foster the specific literacy skill of letter knowledge. However, according to their model, book exposure as an informal source of stimulation should be more relevant for language-related skills such as vocabulary and grammar and hence for later reading comprehension. These assumptions were also confirmed in our study.

Book exposure affects preschool children’s vocabulary and grammar and in turn affects reading comprehension. Furthermore, book exposure is important for developing basic reading skills through grammar. The explanation for the finding that book exposure has an effect on vocabulary, grammar, and reading comprehension seems to come from the complexity of the language the parents use while reading compared to just talking: Mason and Allen (1986), for example, showed that children are exposed to more linguistically complex sentences when someone reads to them. Additionally, the results of Crain-Thoreson, Dahlin, & Powell (2001) indicate that the mean length of utterances is longer when an adult reads to a child. Third, Stanovich and West (1989) showed that the frequency with which a child is read to goes along with more complex oral language use. All in all, children seem to acquire an extended receptive vocabulary and receive a better understanding of the structure of grammar when they are exposed to books. In the same vein, a more sophisticated sentence understanding and better grammar knowledge should lead to better basic reading skills as well. However, one has to keep in mind that basic reading skills were measured through reading speed in the present study.

The present study was able to tie in with the assumptions made by the second trajectory of Snow's (1991, 1999) two-trajectory model of literacy development in school. This second trajectory – reading comprehension – is specifically affected by children's language-related skills of vocabulary and grammar, which in turn are predicted by the informal literacy experience of story book exposure.

In addition to book exposure and as an expansion of Sénéchal et al.'s (1998) model, we considered a second informal source of reading stimulation: The quality of parent-child interactions during book reading. In contrast to our expectations, we did not find direct or indirect effects of interaction quality on reading literacy. However, one has to keep in mind that these effects can be interpreted as effects that are over and above the effects of book exposure and formal instruction as these effects are all modelled simultaneously. In contrast to Lehrl et al. (2012), who found an effect of interaction quality on children's receptive vocabulary in the first year of preschool, the results of the present study did not replicate this effect for children's linguistic skills at the end of preschool. This lack of effect on emergent literacy skills leads to the lack of effect of interaction quality on reading literacy. It seems that the quality of the interaction in a shared book reading situation – measured by the FES – becomes less important when children get older. But because interaction quality affects earlier language development (Lehrl, et al., 2012), it may also boost reading comprehension through autoregressive effects of vocabulary development. Leseman and de Jong (1998) also found a slightly higher effect of literacy opportunity (comparable to our book exposure scale) for the vocabulary of children at the age of 4 (which is the last year of preschool in the Netherlands), than for instruction quality (comparable to our interaction quality scale). As the child's age increases, the overall exposure to books may become more important than the manner in which an adult reads with the child as measured by the FES.

Based on the assumption that later reading skills are determined by the two components (i.e., code-related and language-related skills), parents have several opportunities to support their children: They can assist their children's language competencies through the informal encouragement of interacting with their child while reading (e.g., through asking open-ended questions or providing experiences with books), or they can facilitate code-related skills through the formal teaching of written letters.

Limitations

In spite of the present study's several strengths, such as the longitudinal and multimethod design, the study has some limitations: First, the present study did not consider the effects of the preschool environment. One might argue that parents from a specific advantaged background might select higher quality preschools for their children. As a consequence, improved literacy skills might be traced back to better preschool quality instead of a better HLE. Accordingly, all analyses were also computed while controlling for preschool quality. The effects of the HLE are the same for all outcome variables and can be requested from the corresponding author. Second, book exposure and formal instruction are based on parents' self-reports, which might be affected by social desirability. However, if that was the case, one might expect higher correlations between the two scales. We therefore conclude that social desirability most likely did not cause large measurement error in the present study. Another limitation refers to the fact that because of the design of our study, only children with preschool experience participated in our study. Thus, future research will have to cross-validate the findings by using samples that additionally include children who do not attend preschool.

References

Arbuckle, J. L. (1996). Full information in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and Techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.

Arnold, D. H., & Whitehurst, G. J. (1994). Accelerating language development through picture book reading: A summary of dialogic reading and its effects. In D. K. Dickinson (Ed.), *Bridges to literacy: Children, families, and schools* (pp. 103-128). Oxford, UK: Blackwell Pub.

Bishop, D. (1989). *Test for the Reception of Grammar* (2nd ed.). Manchester: University of Manchester.

Britto, P. R., & Brooks-Gunn, J. (Eds.). (2001). *The role of family literacy environments in promoting young children's emergent literacy skills: New directions for child and adolescent development*. San Francisco: Jossey-Bass.

Burgess, S., Hecht, S., & Lonigan, C. (2002). Relations of the home literacy environment to the development of reading related abilities: a one-year longitudinal study. *Reading Research Quarterly*, *37*(4), 408-426. doi:10.1598/RRQ.37.4.4

Bus, A. G., van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmisson of literacy. *Review of Educational Research*, *65*(1), 1-21. doi:10.3102/00346543065001001

Caldwell, B. M., & Bradley, R. H. (1984). *Home observation for measurement of the environment*. Little Rock: University of Arkansas at Little Rock.

Crain-Thoreson, C., Dahlin, M. P., & Powell, T. A. (2001). Parent-child conversation in three conversational contexts: Variations in style and strategy. In P. R. Britto & J. Brooks-Gunn (Eds.), *The role of family literacy environments in promoting young children's emergent literacy skills: New directions for child and adolescent development* (Vol. 92, pp. 23-38). San Francisco: Jossey-Bass.

de Jong, P. F., & Leseman, P. P. M. (2001). Lasting effects of home literacy on reading achievement in school. *Journal of School Psychology, 39*(5), 389-414. doi:10.1016/S0022-4405(01)00080-2

DeBaryshe, B. D. (1993). Joint picture-book reading correlates of early oral language skill. *Journal of Child Language, 20*(2), 455-462. doi:10.1017/S0305000900008370

Dickinson, D. K. &. Tabors P. O. (1991). Early literacy: Linkages between home, school and literacy achievement at age five. *Journal of Research in Childhood Education, 6*(1), 30-46. doi:10.1080/02568549109594820

Dubowy, M., Ebert, S., Maurice, J. von, & Weinert, S. (2008). Sprachlich-kognitive Kompetenzen beim Eintritt in den Kindergarten: Ein Vergleich von Kindern mit und ohne Migrationshintergrund. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 40*(3), 124-134. doi:10.1026/0049-8637.40.3.124

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test - Revised*. Circle Pines, MN: American Guidance Service.

Ebert, S., Lockl, K., Weinert, S., Anders, Y., Kluczniok, K., & Rossbach, H.-G. (2012). Internal and external influences on vocabulary development in preschool age. *School Effectiveness and School Improvement: An International Journal of Research Policy and Practice*. doi:10.1080/09243453.2012.749791

Evans, D., Shaw, D., & Bell, M. (2000). Home literacy activities and their influence on early literacy skills. *Canadian Journal of Experimental Psychology, 54*(2), 65-75. doi:10.1037/h0087330

Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading & Writing: An Interdisciplinary Journal, 18,* 129-155.

Fox, A. V. (2006). *TROG-D. Test zur Überprüfung des Grammatikverständnisses*. Idstein: Schulz-Kirchner.

Ganzeboom, H. B. G., & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research, 25*(3), 201-239.

Georgiou, G. K., Parrila R., & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology, 100*(3), 566-580. doi:10.1037/0022-0663.100.3.566

Gonzalez, J. E., Taylor, A. B., McCormick, A. S., Villarreal, V., Kim, V., Perez, E., & Darensbourg, A. (2011). Exploring the underlying factor structure of the English version and Spanish translation the "Familia" family literacy inventory: A cautionary tale. *Early Childhood Research Quarterly, 26*(4), 475-483. doi:10.1016/j.ecresq.2010.12.001

Hood, M., Conlon, E., & Andrews, G. (2008). Preschool home literacy practices and children's literacy development: A longitudinal analysis. *Journal of Educational Psychology, 100*(2), 252-271. doi:10.1037/0022-0663.100.2.252

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2,* 127-160.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. doi:10.1080/10705519909540118

Kuger, S., Pflieger, K., & Rossbach, H.-G. (2005). *Familieneinschätzskala (Forschungsversion)*. Bamberg: DFG-Forschergruppe BiKS.

Lehrl, S., Ebert, S., Rossbach, H.-G., & Weinert, S. (2012). Die Bedeutung der familiären Lernumwelt für Vorläufer schriftsprachlicher Kompetenzen im Vorschulalter. *Zeitschrift für Familienforschung, 24*(2), 115-133.

Lenhard, W., & Schneider, W. (2005). *ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässler* (1. Auflage ed.). Göttingen: Hogrefe.

Leppänen, U., Niemi, P., Aunola, K., & Nurmi, J.-E. (2004). Development of reading skills among preschool and primary school pupils. *Reading Research Quarterly, 39*(1), 72-93. doi:10.1598/RRQ.39.1.5

Leseman, P. P. M., & de Jong, P. F. (1998). Home literacy: Opportunity, instruction, cooperation and social-emotional quality predicting early reading achievement. *Reading Research Quarterly, 33*(3), 294-318. doi:10.1598/RRQ.33.3.3

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Lyytinen, P., Laasko, M., & Poikkeus, A. (1998). Parental contributions to child's early language and interest in books. *European Journal of Psychology of Education, 13*(3), 297-308. doi:10.1007/BF03172946

Manolitsis, G., Georgiou, G. K., & Parrila R. (2011). Revisiting the home literacy model of reading development in an orthographically consistent language. *Learning and Instruction, 21*(4), 496-505. doi:10.1016/j.learninstruc.2010.06.005

Manolitsis, G., Georgiou, G., Stephenson, K., & Parrila R. (2009). Beginning to read across languages varying in ortographic consistency: Comparing the effects of non-cognitive and cognitive predictors. *Learning and Instruction, 19*(6), 466-480. doi:10.1016/j.learninstruc.2008.07.003

Mason, J., & Allen, J. B. (1986). A review of emergent literacy with implications for research and practice in reading. *Review of Research in Education, 13*(1), 3-47. doi:10.3102/0091732X013001003

Mayringer, H., & Wimmer, H. (2003). *SLS 1-4 - Salzburger Lese-Screening für die Klassenstufen 1-4*. Bern: Verlag Hans Huber.

Melhuish, E. C., Phan, M. B., Sylva, K., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2008). Effects of the home learning environment and preschool center experience upon literacy and numeracy development in early Primary School. *Journal of Social Issues, 64*(1), 95-114. doi:10.1111/j.1540-4560.2008.00550.x

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6 ed.). Los Angeles, CA: Muthén & Muthén.

Ninio, A. (1983). Joint book reading as a multiple vocabulary acquisition device. *Developmental Psychology, 19*(3), 445-451. doi:10.1037/0012-1649.19.3.445

OECD. (2003). *Literacy skills for the world of tomorrow - further results from Pisa 2000.* Retrieved from http://www.oecd-ilibrary.org/education/literacy-skills-for-the-world-of-tomorrow_9789264102873-en

Payne, A. C., Whitehurst, G. J., & Angell, A. L. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly, 9*(3-4), 427-440. doi:10.1016/0885-2006(94)90018-3

Perfetti, C. A. (1985). *Reading ability*. New York, NY, US: Oxford University Press.

Raikes, H., Pan, B. A., Luze, G., Tamis-LeMonda, C. S., Brooks-Gunn, J., Constantine, J., & et al. (2006). Mother-child bookreading in low-income families: Correlates and outcomes during the first three years of live. *Child Development, 77*(4), 924-953. doi:10.1111/j.1467-8624.2006.00911.x

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

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: evidence, theory, and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of Early Literacy Research* (pp. 97-110). New York: Guilford Press.

Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review, 14*(3), 245-303. doi:10.1006/drev.1994.1010

Sénéchal, M. (2006). Testing the home literacy model: Parent involvement in kindergarten is differentially related to grade 4 reading comprehension, fluency, spelling, and reading for pleasure. *Journal for the Scientific Study of Reading, 10*(1), 59-87.

Sénéchal, M., & LeFevre, J.-A. (2001). Storybook reading and parent teaching: Links to language and literacy development. In P. R. Britto & J. Brooks-Gunn (Eds.), *The role of family literacy environments in promoting young children's emergent literacy skills: New directions for child and adolescent development* (Vol. 92, pp. 39-52). San Francisco: Jossey-Bass.

Sénéchal, M., & LeFevre, J.-A. (2002). Parental involvement in the development of children´s reading skill: A Five-Year Longitudinal Study. *Child Development, 73*(2), 445-460. doi:10.1111/1467-8624.00417

Sénéchal, M., LeFevre, J.-A., Hudson, E., & Lawson, P. (1996). Knowledge of storybooks as a predictor of young children's vocabulary. *Journal of Educational Psychology, 88*(3), 520-536. doi:10.1037/022-0663.88.3.520

Sénéchal, M., LeFevre, J.-A., Thomas, E., & Daley, K. (1998). Differential effects of home literacy experiences on the development of oral and written language. *Reading Research Quarterly, 33*(1), 96-116. doi:10.1598/RRQ.33.1.5

Shanahan, T., & Lonigan, C. J. (2010). The National Early Literacy Panel: A summary of the process and the report. *Educational Researcher, 39*(4), 279-285. doi:10.3102/0013189X10369172

Snow, C. E. (1991). The theoretical basis from relationships between language and literacy in development. *Journal of Research in Childhood Education, 6*(1), 5-10. doi:10.1080/02568549109594981

Snow, C. E. (1999). Social perspectives on the emergence of language. In B. MacWhinney (Ed.), *The emergence of language* (pp. 257-276). Mahwah, NJ: Erlbaum.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, D.C.: National Academy Press.

Son, S.-H., & Morrison, F. J. (2010). The nature and impact of changes in the home learning environment on growth of language and academic skills of preschool children. *Developmental Psychology, 46*(5), 1103-1118. doi:10.1037/a0020065

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24*(4), 402-433.

Stephenson, K. A., Parrila R. K., Georgiou, G. K., & Kirby, J. R. (1998). Effects of home literacy, parents' beliefs, and children's task-focused behaviour on emergent literacy and word reading skills. *Scientific Studies of Reading, 12*(1), 24-50. doi:10.1080/10888430701746864

Storch, A. S., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934-947. doi:10.1037/0012-1649.38.6.934

Teale, W. H., & Sulzby, E. (1989). Emergent literacy: New perspectives. In D. S. Strickland & L. M. Morrow (Eds.), *Emerging literacy: Young children learn to read and write* (pp. 1-15). Newark, DE: International Reading Association.

Torppa, M., Poikkeus, A.-M., Lakos, M.-L., Eklund, K., & Lyytinen, H. (2006). Predicting delayed letter knowledge development and its relation to grade 1 reading achievement among children with and without familial risk for dyslexia. *Developmental Psychology, 42*(6), 1128-1142.

Umek, L. M., Podlesek, A., & Fekonja, U. (2005). Assessing the home literacy environment: Relationships to child language comprehension and expression. *European Journal of Psychological Assessment, 21*(4), 271-281. doi:10.1027/1015-5759.21.4.271

van Steensel, R. (2006). Relations between socio-cultural factors, the home literacy environment and children's literacy development in the first years of primary education. *Journal of Research in Reading, 29*(4), 367-382. doi:10.1111/j.1467-9817.2006.00301.x

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: new evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology, 30*(1), 73-87. doi:10.1037/0012-1649.30.1.73

Weinert, S., & Ebert, S. (in press). Spracherwerb im Vorschulalter: Soziale Disparitäten und Einflussvariablen auf den Grammatikerwerb. *Zeitschrift für Erziehungswissenschaft,*

Weinert, S., Ebert, S., & Dubowy, M. (2010). Kompetenzen und soziale Disparitäten im Vorschulalter. *Zeitschrift für Grundschulforschung, 3*(1), 32-45.

Weinert, S., Ebert, S., Lockl, K., & Kuger, S. (2012). Disparitäten im Wortschatzerwerb: Zum Einfluss des Arbeitsgedächtnisses und der Anregungsqualität in Kindergarten und Familie auf den Erwerb lexikalischen Wissens. *Unterrichtswissenschaft*, *40*(1), 4-25.

Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through book reading. *Developmental Psychology*, *24*(4), 552-559. doi:10.1037/0012-1649.24.4.552

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *69*(3), 848-872. doi:10.1111/j.1467-8624.1998.tb06247.x

Whitehurst, G. J., & Lonigan, C. J. (2003). *Get ready to read! An early literacy manual: Screening tool, activities, and resources*. New York: Pearson Early Learning.

Wygotski, L. S. (1969). *Denken und Sprechen*. Frankfurt a.M.: S. Fischer Verlag.

Zhao, X., Lynch, J. G., JR., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*(2), 197-206. doi:10.1086/651257

Chapter 4

4 Early Literacy Support in Institutional Settings – A Comparison of Quality of Support at the Classroom Level and at the Individual Child Level

Susanne Kuger, Hans-Guenther Rossbach, and Sabine Weinert

Summary

Children's literacy skills and their antecedents start developing very early in life. Next to the family setting, preschools are an important learning context for children prior to school enrollment. Overall, research results point to a strong influence of the quality of stimulation in the classroom on children's literacy development. Yet, a detailed research review reveals that some aspects are more important, whereas others are less important for domain-specific learning support. The research field displays a number of different ways to define educational quality and provides about equally manifold methods to assess it. Most methods that assess educational quality employ observational instruments to measure the

Author Note

Susanne Kuger,
Center for Research on Educational Quality and Evaluation, The German Institute for International Educational Research, Frankfurt, Germany.

Hans-Guenther Rossbach,
Early Childhood Education, University of Bamberg, Germany.

Sabine Weinert,
Developmental Psychology, University of Bamberg, Germany.

This research was supported by grant RO 820/11-1, 11-2, & 11-3 from the German Research Foundation (DFG).

Correspondence concerning this chapter should be addressed to Susanne Kuger, Center for Research on Educational Quality and Evaluation, The German Institute for International Educational Research, Schlossstraße 29, 60486 Frankfurt am Main, Germany. E-mail: kuger@dipf.de

quality of stimulation in the classroom as a whole or the quality of stimulation that is experienced by a single child. The two levels of measurement assess different aspects of educational quality, and they are partially independent of each other, yet both are predictive of children's literacy development. This chapter analyzes single and combined longitudinal relations between quality at the classroom level and at the single child level as well as later reading literacy in a sample of 45 preschool children from the beginning of preschool to the end of the second grade in primary school. Results show that both levels of measurement predict reading literacy in primary school independently of each other but even better when the two measures are combined. Implications for further research and preschool practice are discussed.

Introduction

Literacy competencies in terms of reading and writing abilities are central to children's school success and overall achievement level (Savolainen, Ahonen, Aro, Tolvanen, & Holopainen, 2008). Children's first precursors to later reading skills develop very early in life, which may cause achievement differences in the very first grades of primary school (Duncan et al., 2007). Development in semantics, phonetics, and syntax begins when babies first encounter language and children sometimes recognize letters and "write" symbolic information with their crayons years before they begin formal schooling (Whitehurst & Lonigan, 2001; Stamm, 2003).

For the development of these early literacy skills, early childhood learning conditions are crucial. Learning settings such as the family and non-family care settings offer provisions that can be used to stimulate children's learning prior to formal schooling. Policy makers thereby emphasize the importance of institutional early childhood care, which can foster literacy development for a wide range of students, also reaching out to those children who hail from less stimulating home settings.

Many research studies have demonstrated that the educational quality of institutional early childhood education and care (ECEC) settings is a critical and long-lasting factor in efforts to support children's earlier and later reading achievements and interest (Barnett, Lamy, & Jung, 2005; Camilli, Vargas, Ryan, & Barnett, 2010; Cunningham, 2008; Sammons et al., 2011). Although there is general agreement that high-quality early education and care matters, these current research studies differ with respect to the conceptualization and measurement of educational quality (Halle & Vick, 2007; Pianta & Hamre, 2009). One important difference between studies is the level of assessment of educational quality (Burchinal, 2010). Several ongoing large-scale studies assess educational quality that is offered to a group of children (e.g., Effective Provision of Preschool Education-Study in England, Early Childhood Longitudinal Study-K in the USA), whereas others assess educational quality provided for and experienced by a single child (e.g., NICHD Study of Early Child Care and Youth Development in the USA).

Both assessment methods deliver valuable data on ECEC quality that predict later reading development (e.g., Mashburn et al., 2008; NICHD Early Child Care Research Network, 2006), but the two levels of quality assessment – at the single child level and at the classroom level – do not necessarily capture the same features of educational quality (Layzer, Goodson, & Moss, 1993; Sylva et al., 2007). All children share a certain fraction of common quality experiences, yet at the same time, every child encounters unique situations, activities, and stimulation, which establish a singular experience of educational quality for every single child. So far, there is little information about how the two levels of experience are related to each other and about the degree to which assessments at each level have the power to predict children's later reading achievement (Burchinal, 2010). This chapter focuses on broadening the knowledge and empirical basis of this specific aspect of quality in early childhood education and care. It takes into account the two different levels of quality – the individual child level and the classroom level – and studies their individual and combined explanatory power for later reading achievement in a mid- to long-term view until the end of the second year of primary school.

Literacy in German Early Childhood Institutional Child Care Settings

Child care settings are not a homogeneous group of educational institutions. Their characteristics, educational goals, and realizations depend on national guidelines and policies, cultural understandings of the role of early childcare and educational goals, the overall conditions such as the size and layouts of rooms and furniture, classroom composition, and materials, as well as on the caregivers' understanding and realization of educational goals in the classroom. Some specifics that should be kept in mind when debating about the promotion of early literacy skills in ECEC settings are illustrated in the following. Because this book gives an overview of the results of the German BiKS study, the following section elaborates on specifics about German ECEC settings as they are included in the study.

The Settings

Most child care institutions in Germany are center-based, state subsidized and community- or welfare-led organizations (Rossbach, 2009). So far, only a few but a rising number of for-profit organizations are involved in the German child care market; most organizations are nonprofit or public. Usually the organization, management, and location of ECEC centers are independent from local primary schools with children from an average of three centers enrolled in one primary school. Some German states provide "Vorklassen," a kind of preparatory course in the last year before school entry, and "Eingangsklassen," a special format that combines Grades 1 and 2 to organize a seemingly smoother transition in the years between ECEC and further primary school. Although rather independent from regular primary schools in most regards, the majority of ECEC settings value and emphasize close cooperation with local primary schools.

There is great variation in the duration of a school day. Most settings have traditionally offered child care from about 8 o'clock in the morning until (early) afternoon, but a rising number of mothers in the workforce and a greater demand for extended care provision have led to an extension in the hours of operation from between about 7 o'clock in the morning to 2 to 5 o'clock in the afternoon at most centers. In larger cities or centers that are provided by employers for children of staff members (e.g., in multicorporate enterprises), some child care centers are open from 6 o'clock in the morning until 10 o'clock at night; very few institutions offer overnight services.

Preschool Objectives

The German ECEC system originates from organizations that were first established in the 19th century to provide care and most basic forms of support with regard to questions of health and nutrition for children of working parents. It is in the tradition of this understanding that ECEC organizations used to exclusively depend on and were liable to the social welfare system in Germany and less to the educational system as is the case in many other countries. In recent years, a stronger focus on educational content in the early years of education moved some federal states to involve their educational administration more and more. Although one of the fathers of early childhood pedagogy, Friedrich Froebel, pointed out the importance of education in addition to care issues in the early years, for many decades, most ECEC settings had their primary interest in children's care and supervision (Rossbach, 2008). Thus, for many years, educational goals were located in more general developmental domains such as self-regulation, social behavioral norms, or personal care. Fostering pre-academics and school preparation (i.e., targeting domain-specific educational goals in later school curriculum domains) have therefore been of fluctuating interest. This interest and degree of implementation strongly depended on societal and organizational debates and regained its overall importance only in the last 1 or 2 decades. Beginning in 2002, all federal states prepared and released more or less mandatory curricular guidelines for ECEC institutions, also including pre-academic topics, thus bringing them (back) into the focus of attention in the field.

The Preschool Child

ECEC attendance in Germany is optional and not free of charge. Parents may choose to enroll their child at whichever setting they choose. Mandatory primary school attendance follows different regulations in the different states. In most German states, children are enrolled in primary school around age 6, but begin in ECEC at around the age of 2 or 3 years. Very often, child care prior to preschool is organized in the same settings as preschool for children from the ages of 2 or 3 to 6 years but in different classrooms. Most German ECEC classrooms are attended by age-heterogeneous groups. When the oldest cohort of children leaves the class in summer to transfer to school, new children are integrated in autumn to fill the gap.

Although attendance is optional, the overwhelming majority of German children attend some institutional ECEC setting for more than 1 year. Federal statistics record very high attendance rates (e.g., in 2011, 96.6% of 5-year-olds attended ECEC;

Autorengruppe Bildungsberichterstattung, 2012). Children from disadvantaged backgrounds and children with immigration backgrounds, in particular, display rising but still slightly lower attendance rates.

Educational Quality

Educational studies have focused on the effects of ECEC on child development for a long time, and numerous characteristics and features of child care have been taken into account. In the last 2 decades, a set of measurable characteristics have gained more and more importance in research; these are subsumed under the heading of *educational quality*. This chapter refers to an understanding of educational quality that concentrates on factors that foster healthy overall child development (Bredekamp & Copple, 2006).

A wide number and variety of studies that have focused on the effects of educational quality on child development have been consistent in demonstrating an overall positive effect, yet not all aspects of educational quality have been found to be equally related to different child outcome measures in magnitude. A closer look reveals differential predictive power for various aspects of educational quality for different domains of child development and also for different approaches in their ability to assess educational quality (Anders et al., 2012; Barbarin et al., 2006; Sylva et al., 2006). In alignment with large strands of research on educational effectiveness, common conceptualizations of educational quality have differentiated at least two major aspects: structural background characteristics of the setting and educational processes. Background characteristics have been referred to as "input" with regard to educational situations as they determine the frame and overall conditions of educational interactions. Educational processes in turn involve the child and a teacher, peers, and the physical surroundings such as learning materials. They are conditional on background characteristics and immediately interact with child development. Among educational processes, one can differentiate between different aspects, whereas research has shown that not all aspects support early literacy development equally well. Klieme Lipowsky, Rakotzy, and Ratzka (2006) and Pianta and his colleagues (La Paro, Pianta, & Stuhlman, 2004; Pianta, 2005; Pianta & Hamre, 2009) have distinguished three groups of educational processes that all contribute to process quality: classroom management, climate, and cognitive activation. The first group of processes, classroom management, incorporates aspects related to establishing and maintaining classroom rules and discipline as well as structuring and organizing learning content. Processes referred to as climate help to establish warm and accepting relationships among children and teachers and focus on aspects of respect and emotional support. The last group of processes is aimed at providing highly stimulating learning opportunities that support cognitive development and precursors of academic devolopment. Among the aspects of process quality that aim to stimulate cognitive development, one can further differentiate between educational processes that aim to support the cognitive development of a child in general and educational processes that aim to promote one or more specific developmental domain(s) such as early literacy or numeracy.

Next to this conceptual differentiation of aspects of quality, a differentiation can be made with regard to the level of assessment. Most research studies that predict literacy development and later reading skills on the basis of educational quality in ECEC apply methods to assess educational quality in the preschool class as a whole. A typical approach in these studies is to observe preschool classrooms for some time during average preschool mornings and then to infer the overall educational quality across all conditions and interactions into a single rating of quality in a certain aspect of child care (e.g., overall book use). Research has found meaningful relations between high-quality educational processes in ECEC at the classroom level and children's later reading achievement (e.g., Cunningham, 2010; Dickinson & Caswell, 2007; Sammons et al., 2011; Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2004). A good example is the English longitudinal Effective Provision of Preschool Education (EPPE) study, which found significant long-term effects of quality in ECEC settings on students' achievement up to age 15 (Sammons et al., 2011; Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2010). Still, such a measure is limited in its endeavor to capture individual children's activities and interactions in the classroom and assesses only an abstraction of the variety of interactions in the classroom. Another more finely grained approach for assessing the educational quality that a child experiences during ECEC attendance is to observe this single child's activities and interactions in the classroom as quality indicators (e.g., a child's engagement with books). This approach of assessing educational quality at the single child level leads to a more refined picture.

In turn, the results obtained with this approach are valid only for this child and cannot be generalized to other children in the classroom because every child experiences different activities and interactions with various materials, the teacher, and peers. Studies using process quality at the single child level have also demonstrated predictive power of their quality data for children's later reading development (e.g., Downer & Pianta, 2006).

It can be argued that in a classroom with more overall book use, individual children are also more likely to engage in book use; thus, there is a relation between quality indicators at the classroom level and at the individual child level. Yet research has also shown that even within one and the same classroom and thus among children who experience the same quality at the classroom level, children's profiles of activities vary largely (Sylva et al., 2007), and thus the proportion of shared experiences varies.

Although quality indicators at both levels of assessment – the individual child and classroom levels – have been shown to predict child development, and it is known that both capture different aspects of the quality that a child experiences, thus far, there is little research on how the predictive power of indicators at the two levels are related to each other when studied simultaneously (Burchinal, 2010). Such results could deepen our understanding of the nature of quality at the individual child level and at the classroom level.

Quality of Literacy-Related Processes in Preschool

Although studies do not all apply the same assessment instruments to measure literacy quality and outcome, there seems to be agreement with respect to what is assumed to be at the core of high-quality literacy stimulation in the preschool years. One core principle of educational quality is the developmental appropriateness of all learning opportunities (i.e., personal and physical environments and processes; Bredekamp & Copple, 2006). As illustrated above, most children in Germany spend several years in ECEC settings – as do children in many industrialized countries worldwide (OECD, 2010). During these years, children experience developmental changes in different domains, but very much so in cognitive development and thus also in early literacy, the precursors of later reading and writing skills (Bjorklund, 2004). Developmentally appropriate practice in literacy stimulation implies that educational support is in alignment with this development. Therefore, as children mature and develop cognitively over the course of several years, accompanying high-quality education and care should change in parallel to children's demands and abilities.

High-quality literacy support for a 3-year-old is not necessarily high-quality for a 6-year-old. Whereas familiarizing a child with the habits of book use, the idea of symbolic representation of information in writing, reading to a child, and improving communicative language skills are developmentally appropriate examples of good quality literacy support for a 3-year-old, stimulating the student's awareness of the phonetic structure of language, the rhythm and function of language, letter knowledge, and writing skills might be more appropriate for older children. Such adaptations of domain-specific support that parallel child development can be found across different ECEC curricula (e.g., Neumann, Copple, & Bredekamp, 2000; Neuman & Roskos, 2005).

Research Question

Especially in a domain with large achievement variation at and before school enrollment (e.g., literacy) and in age-heterogeneous classrooms (which even further enlarges achievement variation in comparison to age-homogeneous classrooms), analyzing the difference between process quality at the classroom level and at the individual child level for children's reading literacy development appears to be a highly interesting topic. When caring for an entire class, preschool teachers must address children who are at very different levels of literacy proficiency. Quality at the classroom level therefore needs to take this heterogeneity into account and provide either a large range of possible aspects of support or else provide a level of quality that addresses the abstract commonality of achievement levels, or in other words, the promotion of the "average student." When interacting with an individual child, the teacher can focus much more on this child's current developmental status and adapt possible teaching and interaction strategies to the child. Quality indicators at the classroom level thus should capture the quality that is directed at and provided for an average child or the group of children, whereas quality indicators at the individual child level should differ from that. Thus, quality at the classroom level is assumed to remain rather stable in age-mixed classrooms across the years, whereas quality at the individual child level should display considerable changes as children grow older.

This chapter therefore aims to look at the relations between educational process quality at the single child level and at the class level and at their unique and combined predictive power to explain children's later reading achievement in school (i.e., whether the quality indicators of the two perspectives can be added together or whether they overlap and to what degree they overlap in predicting children's literacy outcomes).

For this question in particular, German preschools are a preferential object of study for two reasons: First, the predominant classroom composition usually includes children within an average age span of 3 to 4 years (ages 2½ up to 6½). Thus, the average achievement range within one classroom is therefore larger than in most other ECEC systems worldwide, and quality aspects at the classroom level and at the individual level should display the largest differences. Second, children remain in the same classroom for several years and in most cases are also taught by the same teacher(s) throughout these years. There is a good chance that the teacher may get to know every child's developmental progress and needs in detail and will adapt his or her teaching strategies and learning opportunities to this knowledge. Therefore, the difference in the effects of the two levels of assessment should be detectable in German settings, perhaps even more distinctly than in other countries' systems.

Method

Adequate study of this research topic necessitates the use of a longitudinal design that includes data on childhood literacy outcomes and educational process quality at both the single child level and at the classroom level.

Sample

The present study used data from a subsample of the longitudinal BiKS-3-10 study. In about half of the preschool classrooms, two different quality assessments were conducted annually on the same day by two different staff members: t_1 in Year 1 (spring 2006), t_2 in Year 2 (spring 2007), and t_3 in Year 3 of children's preschool attendance right before school enrollment (spring 2008). Two children could be observed in each of these classrooms. The BiKS database contains data on later reading achievement in the second grade of primary school for N = 45 children from this subsample (t_4 in spring 2010; only children enrolled in the same school year 2008 with complete observation data at t_1 were included in the analyses). At t_1, during the first assessment of quality indicators, these children had an average age of M = 45.5 months (SD = 2.7). Eight (18%) of the 23 boys and 22 girls had at least one non-German speaking parent and were thus defined as children with an immigration background.

Measures

Early literacy support is related to later reading and writing abilities. The dependent child achievement variable was therefore assessed by a test on reading achievement in primary school. BiKS applied the text comprehension scale of the "Ein Leseverständnistest für Erst- bis Sechstklässler" (ELFE 1–6; Lenhard & Schneider, 2009), a test of reading comprehension for first to sixth graders. This subtest of about 7-min duration applies 20 multiple-choice items testing for students' ability to pick out relevant information from a short text and to draw inferences from this information. The internal consistency (Cronbach's alpha) of this scale for the relevant measurement point in Grade 2 is high (α = .94). The children's language development was assessed annually in terms of receptive vocabulary with a German version of the Peabody Picture Vocabulary Test (PPVT-R; Dunn & Dunn, 1981). Language testing took place about three months prior to the quality observations each, that is, the first assessment wave of language proficiency was winter 2005/06, and preschool quality was observed in spring 2006 (parallel for later assessment points; for further reading on the BiKS-design, see Lorenz, Schmitt, Lehrl, Mudiappa, & Roßbach, chapter 2, this volume).

The BiKS study includes questionnaires for preschool teachers and parents as well as observational measures (cf. Lorenz et al., chapter 2, this volume). Process quality at the classroom level and at the individual child level was assessed through live rating observations on the same preschool morning. The two assessments were conducted by two different observers (after several days of schooling, observers had to reach an 80% agreement with the training research staff on all observation measures in order to be part of the field staff). Quality at the classroom level was assessed using the German versions of the ECERS-R (Harms, Clifford, & Cryer, 1998) and ECERS-E (Sylva, Siraj-Blatchford, & Taggart, 2003) rating scales. The two instruments cover a wide range of education and care topics in early childhood settings. Indicators are scored on a 7-point rating scale (1 = *lower quality* to 7 = *better quality*). An indicator of quality of *literacy and language support at the classroom level* (LCL) was created across the two instruments by computing the mean score of the following items: books and pictures, encouraging children to communicate, informal use of language, environmental print: letters and words, book and literacy areas, adult reading with the children, sounds in words, emergent writing/mark making, and talking and listening (internal consistency Cronbach's alpha: t_1 = .74; t_2 = .78; t_3 = .72).

Quality at the individual child level was assessed using a newly developed tool. This target child observation is related to earlier instruments of individual child observations such as the ORCE (National Institute of Child Health and Human Development, 1996) and the OAP (Lera & Palacios, 1995) but advances these earlier instruments by adding a focus on the quality of domain-specific activities related to literacy and numeracy, for example. The instrument allows for three cycles of 20-min observations across an average morning. In every cycle, observers note the quality of education and care for a number of different global and domain-specific aspects of process quality. Because definitions of early literacy vary widely, this chapter includes two versions of quality of literacy stimulation at the *individual child level*: one follows a more *narrow definition of early literacy*, which is mainly focused on support in code-related skills (mean of ratings in use of letters, [pre-]reading and pretending to read, and [pre-]writing and pretending to write), therefore called *literacy support* (NLIL; internal consistency Cronbach's alpha t_1 = .68; t_2 = .53; t_3 = .67); the indicator for a *broader definition of literacy* includes ratings on these three items and in addition on the item "use of questions in interactions". Thus, the second indicator is less specific, also covering topics of a more general cognitive and language support, and is therefore called *literacy and language support* (BLIL; the broadness of the indicator results in low internal consistency: Cronbach's alpha t_1 = .32; t_2 = .45; t_3 = .39). Every item represents the mean of three periods of observation across a typical preschool morning.

Questionnaires for parents were applied to assess the children's family background characteristics such as their immigration background and the families' socio-economic status (SES), which was measured using the highest value of both parents' international socio-economic index (Ganzeboom, de Graaf, Treiman, & de Leeuw, 1992; HISEI).

Analyses and Procedure

As in most studies with repeated measurements, some missing data were to be found in the data. Missing data analyses suggested that they were missing at random. The literature in this case advises that missing data be taken into account as such rather than reducing the sample size via listwise deletion (Lüdtke, Robitzsch, Trautwein, & Köller, 2007). The sample therefore represents all students who were included in the subsample of parallel quality measurement and for whom there was achievement data for the second-grade reading test (sample as described above). The data were analyzed using the software package MPlus 5 (Muthén & Muthén, 2008), which applies the full maximum likelihood (FIML) approach to account for missing data and also takes into account the clustered sample structure (up to two children per preschool).

In a first step, quality indicators were correlated with each other in order to analyze the degree of relatedness among quality indicators, to determine the degree to which the two levels of assessment were related to each other, and whether the relation changed over the course of three consecutive preschool years. As the children developed, we expected quality measures at the single child level to change, whereas quality at the classroom level was expected to remain rather stable. Next, quality indicators were correlated with children's vocabulary development to study the pattern of relatedness of literacy quality to children's developmental path and whether quality at the individual child level was adapted to the children's progress. Finally, both quality indicators were studied in their individual and combined relation to children's later reading achievement in multiple regression analyses controlling for the most relevant child background variables (age at assessment of reading achievement t_4 in grade 2, SES, immigration background, and vocabulary status in the first year of ECEC at the age of 3 years).

Results

At t_1, the children's parents' average HISEI was 51.4 (SD = 16.5) and children's vocabulary knowledge in this first year of preschool averaged 27.1 words on the PPVT (SD = 11.8; Year 2: M = 48.49, SD = 14.1; Year 3: M = 74.6, SD = 17.1). Student's reading achievement in the second grade displayed an average test score of 9.6 (SD = 4.4) correct answers for this subsample of children who were then 97.5 months old (SD = 4.4; ≈ 9 years 2 months). Descriptive results of both indicators of process quality are indicated in Table .

Table 1. Descriptive Results of Quality Measures

	t_1 spring 2006 *M (SD)*	t_2 spring 2007 *M (SD)*	t_3 spring 2008 *M (SD)*
Literacy support at the individual child level (NLIL)	1.1 (0.14)	1.1 (0.19)	1.2 (0.31)
Literacy and language support at the individual child level (BLIL)	1.5 (0.20)	1.6 (0.22)	1.7 (0.33)
Literacy and language support at the classroom level (LCL)	3.9 (0.71)	4.2 (0.81)	3.9 (1.00)

Note. All indicators range from a scale minimum of 1 to a scale maximum of 7.

Descriptive results point to the lack of emphasis that was placed on very early literacy instruction in German preschools. Overall provision of literacy and language support at the classroom level (LCL) reached a level of medium quality. Comparing the two indicators for individual children's experiences, the data indicated that this was largely due to more overall language stimulation and not to literacy support in the narrow sense. Although quality at the individual child level was low for both indicators and all measurement points, the quality of code-related literacy promotion at the individual level (NLIL) was even lower than the broader indicator of literacy and language support (BLIL). Both were lowest in the first year of preschool and increased only marginally while vocabulary changed significantly (Ebert et al., 2012; Weinert, Ebert, Lockl, & Kuger, 2012). Conclusions drawn from further analyses thus need to take into account these floor effects (and the low variability in these measures).

Table 2. Bivariate Correlations of the Quality Measure at the Classroom Level with Measures at the Individual Level

		Literacy and language support at the classroom level (LCL)		
		t1	t2	t3
Literacy support at the individual child level (NLIL)	t1	.01	.12	.06
	t2	.06	.21	.40***
	t3	-.14	.09	.00
Literacy and language support at the individual child level (BLIL)	t1	.17	.34	.17
	t2	.17	.32**	.34**
	t3	-.16	-.05	.51***

+ $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Bivariate correlations between indicators at the individual child level and at the classroom level displayed very small relations between levels of assessment in the first year of ECEC (all *ns*). Comparing the upper and the lower halves of Table 2, it became evident that relations between the classroom level (LCL) and the conceptually broader indicator at the individual child level (BLIL) were slightly stronger than those between LCL and the narrower indicator, NLIL (average $r_{\text{BLIL, LCL}} = .22$; average $r_{\text{NLIL, LCL}} = .09$; one exception from this bias is $r_{\text{NLILt3, LCLt2}} = .40$). This pattern was found throughout the years of ECEC attendance. The overall level of relations rose in Year 2 and Year 3 in particular for the broader indicator at the individual child level, BLIL ($r_{t1} = .17$; $r_{t2} = .32$**; $r_{t3} = .51$***). Taking into account the items included in the scales as enumerated in Section 6.2 (Measures), it could be expected that literacy at the classroom level follows a broader definition of literacy including a wider variety of aspects as did the broader definition of literacy and language at the individual child level. But literacy and language promotion at the classroom level also seemed to be oriented towards an average standard of literacy process quality that was usually experienced by children in their second and third or last year of ECEC rather than in their first year of ECEC. This finding is in contradiction to the usually implicit assumption that the ECERS scales cover educational quality equally well and imply the same meaning for all children in ECEC. Given these results, ECERS values might have a different meaning for the stimulation of 3-year-olds, 4-, 5-, or 6-year-olds.

Besides this description of patterns of relations among different indicators of educational quality, this chapter seeks to research the relative predictive power of different indicators for children's achievement. The study included the PPVT as a measure of the children's receptive vocabulary. Table 3 displays correlations between language outcomes (vocabulary in Year 1, Year 2, and Year 3 of ECEC attendance and reading achievement in the second grade of primary school) and indicators of process quality.

Table 3. Bivariate Correlations between Language and Reading Outcomes and Quality Measures during the Years of ECEC attendance

		Vocabulary in preschool year 1 of ECEC	Vocabulary in preschool year 2 of ECEC	Vocabulary in preschool year 3 of ECEC	Reading achievement grade 2 in primary school
Literacy support at the individual child level (NLIL)	t1	-.17	-.01	-.19	.13
	t2	.14	.18	.06	-.15
	t3	-.42 *	-.39 ***	.00	-.41 ***
Literacy and language support at the individual child level (BLIL)	t1	.05	.07	-.09	.27 *
	t2	.05	.16	.15	-.17
	t3	-.08	-.23	-.04	-.21
Literacy and language support at the classroom level (LCL)	t1	.28	.39 **	.26 +	.43 **
	t2	.04	-.01	-.05	-.05
	t3	.25 +	.13	.02	-.09
Vocabulary in year 1 of ECEC	year 1				.48 ***
	year 2				.55 ***
	year 3				.27 *

+ $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table displays very low relations between support at the individual child level and children's language proficiency in terms of receptive vocabulary during the years of ECEC attendance. Most correlations were not significant and many were close to zero. The only practically relevant relations between support and language proficiency were observed in Year 3. Literacy support at the individual child level (NLIL) in the last year of ECEC displayed significant negative moderately sized relations with children's vocabulary results in earlier years (vocabulary in Year 1: $r = -.42$*; vocabulary in Year 2: $r = -.39$***). This result points to a compensatory reaction of ECEC settings in the last year before school enrollment to some children's earlier low language proficiency. Results for the broader indicator of individual support (BLIL) were similar, but far less strong and not significant. Comparing the pattern of relations, the settings' efforts to provide support in the last year of ECEC seemed to be concentrated mainly on code-related literacy promotion (use of letters, pre-writing, and pre-reading) and not on the broader range of language and literacy support additionally including conversational skills and questions that are cognitively stimulating. Current language abilities in Year 3 seemed irrelevant for the support provided. Support at the classroom level, on the other hand, displayed a tendency to be better for children with better language proficiency during ECEC attendance (cf. Table 3).

Relations of support and receptive vocabulary knowledge throughout the ECEC years with later reading achievement supported two points of interpretation in particular: (a) The significant negative correlation of medium size between NLIL and later reading achievement supported the assumption of a compensatory reaction to earlier low language proficiency in Year 3 (cf. Table 3) and at the same time indicated that these measures of treatment might have only a small impact on children's further development: Children's vocabulary scores in the ECEC years were significantly related to later reading achievement (Year 1 vocabulary with second-grade reading: $r = .48$***; Year 2: $r = .55$***; Year 3: $r = .27$*). Children with lower vocabulary knowledge in the earlier years received better individual literacy support in Year 3 of ECEC, whereas children with better vocabulary knowledge in Year 1 experienced less support (see above Year 1 vocabulary with Year 3 NLIL: $r = -.42$*; Year 2 vocabulary with Year 3 NLIL: r = -.39***). But such slightly better support in the last year before school enrollment was significantly related to lower reading achievement in the second grade (Year 3 NLIL with second grade reading: $r = -.41$***). Students did not seem to profit very much from these measures of support. (b) At the same time, very early (Year 1) promotion of a broader understanding of literacy and language support was significantly and positively related to later reading achievement (Year 1 BLIL with second-grade reading: $r = .27$*; Year 1 LCL with second-grade reading $r = .43$**) but not with synchronous vocabulary knowledge. A broader combined stimulation of code-related and communication skills seemed to be more beneficial (in terms of longitudinally positive relations, but perhaps not purely causal effects) for children's later reading ability. This long-term positive relation of support and child outcome across 4 years of child development (the first 3 years of preschool to second grade in primary school) was not replicated using later measures of child support.

Because educational quality is assumed to impact students' outcome in the long run and because overall the strongest relations of support with later reading achievement were found for data from Year 1 of ECEC attendance, these early measures were used to further analyze their individual and combined relations beyond bivariate correlations in multiple regression analyses. Vocabulary in Year 1 of ECEC was also strongly related to later reading achievement and related to some quality measures in Year 1 (i.e., significantly related to quality measures at the classroom level). Further analyses therefore controlled for early vocabulary knowledge. The multivariate analyses were conducted in parallel for both conceptualizations: the narrow and broad definitions of literacy.

Table 4. Multivariate Linear Regression of Reading Achievement on Educational Quality

	Model 0	Model 1	Model 2a	Model 3a	Model 2b	Model 3b
Migration background (ref. immigration background)	.11	.19*	.11	.18*	.10	.18+
SES	.19	.23*	.17	.22*	.22+	.25*
Vocabulary preschool Year 1	.30+	.18	.36*	.23+	.32*	.24+
Age Grade 2	.22	.29*	.17	.26+	.15	.21+
Literacy and language support at the classroom level (LCL)		.44**		.42***		40**
Literacy support at the individual level (NLIL)			.18	.14		
Language and literacy support at the individual level (BLIL)					.29*	.20*
R^2	.26+	.43**	.29*	.45**	.33**	.47***
N	45	45	45	45	45	45

+ $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$.

As expected regarding the construction of indicators and the bivariate correlations reported above, the two versions of analyses led to a parallel pattern of results. The background model (Model 0) explained 26% of the variance between students, but was not significant. After controlling for family SES, students' immigration background, and age at assessment of reading achievement, vocabulary in the first year of ECEC displayed the largest relation to students' reading achievement in the second grade.

The amount of explained variance increased substantially when educational quality in literacy support at the classroom level was included in the analyses, and it was the most important predictor in Model 1. It should be noted that after literacy support at the classroom level was included in the model, students' background characteristics became significant. For all models, literacy support at the classroom level remained the most important predictor for later reading achievement.

Examining the two models that included literacy support at the individual child level but not at the classroom level (Models 2a and 2b), the most obvious change from the background model was that only the broader indicator of literacy and language support at the individual child level contributed significantly to the overall model. Compared to the background model, the more narrow understanding of literacy support (NLIL; Model 2a) increased the overall amount of explained variance by only 3% (ΔR^2: *ns*), whereas the broader indicator of literacy and language support (BLIL; Model 2b) added 7% of explained variance (ΔR^2: $p < .05$). The indicator of a broader understanding of literacy support in preschool predicted later reading achievement almost as well as earlier vocabulary knowledge did.

Models 3a and 3b both incorporated indicators of literacy support at the classroom level and at the individual child level and as expected, explained the largest amount of variance. In Model 3b, literacy and language support at the individual child level contributed significantly to the overall explanatory power, whereas only the families' SES retained its significance from the background model. This model was also the most predictive, explaining almost half the variance in later reading achievement. Finally, the models holding only literacy support on classroom level should be compared to those that additionally include an indicator at individual child level (models 1 and 3a for NLIL, models 1 and 3b for BLIL). Change in overall R^2 was very small and not significant for the narrow definition of literacy support on individual child level (NLIL; $\Delta R^2 = .02$; *ns*), and slightly bigger and tending to significance for the broader indicator of language and literacy support (BLIL; $\Delta R^2 = .04$; $p < .1$).

Summarizing the results from the regression analyses, the study showed that process quality at the individual child level for literacy and for a broader indicator of literacy and language support both positively contributed to a background model in explaining later reading achievement, but the very narrow understanding of code-related literacy support at the age of 3 did not contribute significantly. Literacy and language support at the classroom level, on the other hand, had a very strong relation to a later text comprehension outcome. Combining quality at the individual child level with quality at the classroom level led to an even better prediction of later achievement. This came along with two patterns of results: The indicator of quality at the classroom level remained the strongest predictor throughout all models, and its impact was reduced only slightly after educational quality at the individual child level was included; simultaneously, the impact of process quality at the individual child level was reduced somewhat more strongly when literacy support at the classroom level was included, and only the broader conceptualization of literacy and language support reached significance after controlling for literacy support at the classroom level. Thus both levels of quality assessment contribute individual shares to the prediction of later reading achievement but this prediction is better for broader concepts of literacy support which not only focus on code-related skills but more overall language support in early ages as well.

Discussion

The study included a small subsample of children from the BiKS-3-10 study for which complete data on reading achievement in second grade of primary school is available and educational process quality in literacy and language support in the first year of preschool was measured at two levels of assessment: individual child level and classroom level.

Results first of all point to the low level of literacy support in German ECEC during the years of study (2006-2008). Not so much in terms of the overall level of support and presence of literacy and language in the classroom, but regarding individual children's experiences and the degree of literacy and language support that aims to promote individual children's development. Educational quality at individual child level is very low. Since the observational instrument used to assess educational quality at individual child level was newly developed for the purposes of the BiKS study, this result could be caused by different reasons. Floor effects could be a purely methodological effect of an overly ambitious scale, i. e. the instrument could demand too high standards for at least minimum quality ratings. Yet the scale was developed on the basis of international standards of good practice and other instruments available in the field such as the ELLCO (Smith, Brady, & Anastasopoulos, 2008) and should thus be valid. Results more probably reflect real low levels of individualized literacy support in German preschools in the years of assessment. Nevertheless these floor effects should be kept in mind in further interpretation of the results, as they might explain an overall low level of relatedness to other indicators.

Although process quality of literacy and language support displays medium values at most, indicators at both levels of assessment were related to later reading achievement from a long-term perspective across the 4-year time span of the study. Later quality measures were less strongly related to reading achievement in the second grade. Those children who experienced good quality at the very beginning of their years of ECEC attendance displayed better reading achievement later in primary school. This result is in line with other international research. Results from the EPPE study in England (Sammons et al., 2004; Sammons et al., 2011; Sylva et al., 2010) had shown that ECEC quality measured at the age of 3 had a long-lasting effect on different cognitive and socio-emotional domains of child development up to the second grade in primary school and far beyond. The EPPE study missed later assessments of quality throughout the years of ECEC as they were included in BiKS. Whether process quality unfolds its maximum “impact” on child development in the long run or whether early experiences of quality are most critical for later achievement (as the results of the current study indicate) will have to be determined by future analyses that also include data from even later measurement points of the BiKS study.

The differences found between the narrower and broader definitions of literacy support at the individual child level are important to mention here. Whereas very early literacy support in a broader sense was positively related to later reading achievement, support as more narrowly defined was not positively related to later reading achievement. Moreover, children with lower language proficiency in the early years experienced better literacy support as narrowly defined in the later years of ECEC. This can be interpreted as a compensatory reaction in classrooms to support low achievers prior to school enrollment. If this endeavor was successful, results should display positive relations of later support in ECEC with reading achievement in grade 2. But better late literacy support as narrowly defined (i.e., possibly compensatory endeavors) is related to lower reading achievement in the second grade just as lower early vocabulary skills are. It can be concluded that if settings have the goal of compensating for the low language achievement of some children, late literacy support in a very narrow sense cannot do the job well enough by itself.

A comparison of the quality of literacy support at the classroom level and at the individual child level displayed larger relations between the indicator at the classroom level with a broader understanding of literacy and language support at the individual child level than with a more narrow definition of mainly code-related literacy support. This is most probably due to the fact that the indicator at the classroom level itself made use of a broader definition that included, for example, overall book use and language support. Thus, the difference in relations points to conceptual relatedness and differences but also to a shared concept of quality that is independent of assessment level.

Overall, it seems that broader support (i.e., a combination of promotion in literacy and language domains) is more beneficial for later reading achievement than a more narrowly focused promotion of code-related skills only. Given that reading acquisition and achievement is determined by numerous factors, going far beyond letter knowledge, recoding, and writing skills – which were included in the narrow realization of individual literacy support – the results of this study once again underline the importance of support across a broader range of domains. The broader indicators at the classroom level and at the individual child level in this study included aspects such as asking cognitively stimulating questions, using language to support cognitive development, or engaging in longer conversations with children. Besides stimulating language alone, these also promote children's overall cognitive and meta-cognitive development and thus contribute to a number of different developmental domains, which in turn all have a share in reading acquisition and later achievement.

As a limitation, it should be noted that low relations of the narrow realization of literacy support at the individual child level and reading achievement could also be due to the small amount of variance in the quality indicator caused by a floor effect. However, a similar floor effect was also observed in the broader indicator of individual support, which did not prevent this measure from displaying a stronger relation to reading achievement.

The results of the multiple regression analyses additionally supported the existence of shared and non-shared components of process quality in the indicators of process quality at the different levels of assessment. Indicators at both levels predicted later reading achievement independently from each other, but also shared a combined understanding of educational quality. For this study, quality at the classroom level was much more important than quality assessed at the individual child level. At least two different explanations for this finding should be discussed. One is that this is due to a methodological issue. After all, regarding internal consistency, variance, and skewness, the measure at the classroom level delivered better data than did the indicators at the individual child level. Another possible interpretation could be that educational process quality that is shared among children in the classroom has more impact than quality experienced by just an individual child. Process quality at the classroom level interacts directly with a child, but may furthermore interact indirectly through the child's peers, who also profit from quality in this classroom and in turn stimulate language and literacy development in the target child. An analysis that includes language proficiency and the development of all students in the classroom could further illuminate this line of argument. Nevertheless, quality at the level of assessment of individual children could additionally contribute to the prediction of later reading achievement and could thus conceptually provide information about educational quality that cannot be covered by indicators at the classroom level of assessment.

Further details about the nature of shared and non-shared components of process quality cannot be analyzed in this study because of several limitations. First, the sample was rather small so that it was not possible to develop models to test the impact of a wider variety of children's background characteristics or to test for differential results through interaction effects. A replication of the study with a larger sample could therefore add valuable information about the differences between the results and the concepts of process quality at the individual child level and at the classroom level. Second, knowing about the low level of quality of literacy (and language) support at the individual child level, it might be feasible to include lower level quality indicators in the description of the instrument to obtain a better differentiation among preschools in the lower quality range (which is true for most settings). Results might profit from a larger variance. Still, it must be underlined that currently the lowest level of quality described in the instrument constitutes a very low level of stimulation: the item "writing and precursors of writing," for example, should be given a rating of "1" (scale minimum) if the teacher does not help the child to write anything, the child is not encouraged to write anything, the child is not given support for writing spontaneously (e.g., praise), or if the child does not experience any writing in the classroom. It might be advantageous for research purposes but would be difficult and questionable for practical reasons to find descriptors for even lower levels of quality of early literacy support.

References

Anders, Y., Rossbach, H.-G., Weinert, S., Ebert, S., Kuger, S., Schmidt, S., & von Maurice, J. (2012). Home and preschool learning environments and their relations to the development of early numeracy skills. *Early Childhood Research Quarterly, 27,* 231-244. doi:10.1016/j.ecresq.2011.08.003

Autorengruppe Bildungsberichterstattung. (2012). *Bildung in Deutschland 2012. Ein indikatorengestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf.* Bielefeld: W. Bertelsmann Verlag.

Barbarin, O., Bryant, D., McCandies, T., Burchinal, M., Early, D., Clifford, R., & ... Howes, C. (2006). Children enrolled in public pre-K: The relation of family life, neighborhood quality, and socioeconomic resources to early competence. *American Journal Of Orthopsychiatry, 76*(2), 265-276. doi:10.1037/0002-9432.76.2.265

Barnett, W. S., Lamy, C., & Jung, K. (2005). *The Effects of State Prekindergarten Program on Young Children's School Readiness in Five States*. Rutgers University, National Institute for Early Education Research.

Bjorklund, D. F. (2004). *Children's thinking: Cognitive development and individual differences*. Belmont, CA: Wadsworth Publishing.

Bredekamp, S., & Copple, C. (2006). *Basics of developmentally appropriate practice.* Washington, DC: National Association for the Education of Young Children.

Burchinal, M. (2010). *Differentiating among measures of quality: Key characteristics and their coverage in existing measures. (Research-to-Policy, Research-to-Practice Brief OPRE 2011-10b)*. Washington, DC: U.S. Administration for Children and Families, Office of Planning, Research and Evaluation. Retrieved 2013-04-24 from http://www.acf.hhs.gov/sites/default/files/opre/differ_measures.pdf.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W.S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*(3), 579-620.

Cunningham, D. (2010). Relating Preschool Quality to Children's Literacy Development. *Early Childhood Education Journal, 37*(6), 501-507. doi:10.1007/s10643-009-0370-8

Cunningham, D. D. (2008). Literacy environment quality in preschool and children's attitudes toward reading and writing. *Literacy Teaching and Learning, 12*(2), 19-36.

Dickinson, D. K., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly, 22*(2), 243-260. doi:10.1016/j.ecresq.2007.03.001

Downer, J. T., & Pianta, R. C. (2006). Academic and Cognitive Functioning in First Grade: Associations with Earlier Home and Child Care Predictors and with Concurrent Home and Classroom Experiences. *School Psychology Review, 35*(1), 11-30.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428-1446. doi:10.1037/0012-1649.43.6.1428

Dunn, L. M., & Dunn, L. M. (1981). *Examiner's manual Peabody Picture Vocabulary Test - Revised Edition*. Circle Pines: American Guidance Service.

Ebert, S., Lockl, K., Weinert, S., Anders, Y., Kluczniok, K., & Rossbach, G. (2012). Internal and external influences on vocabulary development in preschool age. *School Effectiveness School and School Improvement: An International Journal of Research, Policy and Practice*. doi:10.1080/09243453.2012.749791

Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research 21*(1), 1-56.

Halle, T., & Vick, J. E. (2007). *Quality in Early Childhood Care and Education Settings: A Compendium of Measures*. Washington, DC: Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale.* Revised Edition. New York: Teachers College Press.

Klieme, E., Lipowsky, F., Rakoczy, R., & Ratzka, N. (2006). Qualitatsdimensionen und Wirksamkeit von Mathematikunterricht. In M. Prenzel & L. Allolio-Nacke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule* (pp. 127-146). Münster: Waxmann.

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The elementary school journal, 104,* 409-426.

Layzer, J., Goodson, B. D., & Moss, M. (1993). *Observational study of early childhood programs: Final report volume I: Life in preschool*. Cambridge Mass.: Abt. Associates.

Lenhard, W., & Schneider, W. (2006). *ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe.

Lera, M.-J., & Palacios, J. (1995). *OAP: Observación de Actividades Preescolares* (unpublished document). Departamento de Psicología Evolutiva y de la Educación, Universidad de Sevilla.

Lüdtke O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau, 58*(2), 103-117.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., ... Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic language and social skills. *Child Development, 79*(3), 732-749. doi: 10.1111/j.1467-8624.2008.01154.x

Muthén, L. K., & Muthén, B. O. (2008). *Mplus user's guide* (Version 5). Los Angeles: Muthén & Muthén.

National Institute of Child Health and Human Development (NICHD) (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 11,* 269-306.

Neuman, S. B., & Roskos, K. (2005). The state of state pre-kindergarten standards. *Early Childhood Research Quarterly, 20*(2), 125-145. doi:10.1016/j.ecresq.2005.04.010

Neuman, S. B., Copple, C., & Bredekamp, S. (2000). *Learning to read and write: Developmentally appropriate practices for young children*. Washington, DC US: National Association for the Education of Young Children.

NICHD Early Child Care Research Network. (2006). Child Care Effect Sizes for the NICHD Study of Early Child Care and Youth Development. *American Psychologist, 61*, 99-116. doi: 10.1111/j.1467-8624.2010.01431.x

OECD (2010). *PISA 2009 Results: Overcoming Social Background: Equity in Learning Opportunities and Outcomes (Volume II)*. PISA, OECD Publishing. doi: 10.1787/9789264091504-en

Pianta, R. C., & Hamre, B. K. (2009): Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educational Researcher, 38*(2), 109-119.

Pianta, R. C., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child–teacher interactions? *Applied Developmental Science, 9*(3), 144-159. doi:10.1207/s1532480xads0903_2

Rossbach, H. G. (2008). Vorschulische Erziehung. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.): *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (pp. 283-323). Reinbek: Rowohlt.

Rossbach, H.G. (2009). The German Educational System for Children from 3 to 10 Years Old. In R. M. Clifford & G. M. Crawford (Eds.). *Beginning School. U.S. Policies in International Perspective* (pp. 53-67). New York: Teachers College Press.

Sammons, P., Sylva, K., Melhuish, E. C., Siraj-Blatchford, I., Taggart, B., Elliot, K., & Marsh A. (2004). *The Effective Provision of Pre-School Education (EPPE) Project: Technical Paper 11 – The continuing effects of pre-school education at age 7 years*. London: DfES / Institute of Education, University of London.

Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., Toth, K., ... Smees, R. (2011). *Effective Pre-School, Primary and Secondary Education Project (EPPSE 3-14): Influences on students' attainment and progress in Key Stage 3: Academic outcomes in English, maths and science in Year 9.* London: Institute of Education, University of London / Department for Education.

Savolainen, H., Ahonen, T., Aro, M., Tolvanen, A., & Holopainen, L. (2008). Reading comprehension, word reading and spelling as predictors of school achievement and choice of secondary education. *Learning & Instruction, 18*(2), 201-210. doi:10.1016/j.learninstruc.2007.09.017

Smith, M. W., Brady, J. P., & Anastasopoulos, L. (2008). *Early Language and Literacy Classroom Observation Tool (ELLCO Pre-K)*. Baltimore: Brooks Publishing.

Stamm, M. (2003). Looking at long-term Effects of early reading and numeracy ability: a glance at the phenomenon of giftedness. *Gifted and Talented International, 18*(1), 7-16.

Sylva, K., Melhuish, E. C., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004),The Effective Provision of Pre-School Education (EPPE) Project: *Technical Paper 12 – The Final Report: Effective Pre-School Education*. London: DfES / Institute of Education, University of London.

Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing quality in the early childhood rating scale extensions (ECERS-E)*. Stoke On Trent: Trentham Books Ltd.

Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsika, V. (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly, 21*(1), 76-92. doi:10.1016/j.ecresq.2006.01.003

Sylva, K., Taggart, B., Siraj-Blatchford, I., Totsika, V., Ereky-Stevens, K., Gilden, R., & Bell, D. (2007). Curricular quality and day-to-day learning activities in pre-school. *International Journal of Early Years Education, 15*(1), 49-65. doi:10.1080/09669760601106968

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2010). *Early Childhood Matters: Evidence from the Effective Pre-school and Primary Education Project*. Oxford: Routledge.

Weinert, S., Ebert, S., Lockl, K., & Kuger, S. (2012). Disparitäten im Wortschatzerwerb: Zum Einfluss des Arbeitsgedächtnisses und der Anregungsqualität in Kindergarten und Familie auf den Erwerb lexikalischen Wissens. *Unterrichtswissenschaft, 40,* 4-25.

Whitehurst, G., & Lonigan, C. (2001). Development from pre-readers to readers. In S. Neuman & D. Dickinson (Eds.), *Handbook of Early Literacy Development* (pp. 11-29). New York: Guilford.

Chapter 5

5 Predicting Reading Literacy in Primary School: The Contribution of Various Language Indicators in Preschool

Susanne Ebert and Sabine Weinert

Summary

Although children's language competencies in preschool are known to be important predictors of reading literacy, the nature of the relation between early language and later reading literacy is still under debate. This is presumably due to the multicomponential nature of language as well as of reading literacy. In this chapter, we begin with a brief overview of theoretical assumptions and empirical results regarding how various facets of language are connected to reading literacy. However, the majority of the existing empirical studies do not clearly differentiate between various aspects of the individual's language and reading literacy and often consider only single aspects of language and/or reading. Therefore, data from the longitudinal BiKS-3-10 study were used to more directly compare the impacts of various indicators of early language competencies on different aspects of reading literacy. Specifically, we considered the importance of (a) phonological information processing skills (phonological working memory, speed of access to long-term

Author Note

Susanne Ebert,
Department of Developmental Psychology, University of Bamberg, Germany.

Sabine Weinert,
Department of Developmental Psychology, University of Bamberg, Germany.

This research was supported by grant WE 1478/4-1, WE 1478/4-2, & WE 1478/4-3 from the German Research Foundation (DFG).

Correspondence concerning this chapter should be addressed to Susanne Ebert, Department of Developmental Psychology, University of Bamberg, Markusplatz 3, 96045 Bamberg, Germany. E-mail: susanne.ebert@uni-bamberg.de

memory) and (b) linguistic abilities in the sense of language components (vocabulary, grammar) as well as more integrative language competencies (story reproduction and comprehension, sentence reproduction) in preschool on (c) rather basic reading skills and (d) reading comprehension in the second year of primary school. In contrast to many other studies, the BIKS-3-10 study assessed various potentially relevant language predictors of reading, and in addition, this assessment was conducted at an early age of about 4 years. In particular, we examined whether early linguistic abilities in the sense of vocabulary and grammar would be – as often assumed – more strongly associated with reading comprehension, whereas early phonological processing skills would be more strongly associated with more basic aspects of reading development such as reading speed. Additionally, we asked whether integrative language competencies (story reproduction and comprehension, sentence reproduction) would be more predictive of early reading comprehension than measures of linguistic abilities in the sense of language components (i.e., vocabulary, grammar). The results of the BiKS-3-10 study are discussed with regard to different theories and assumptions about the ways in which language is predictive of reading literacy development.

Reading literacy is – undoubtedly – a key competence in modern societies. Interindividual differences and individual deficits in reading abilities tend to show up rather early in school and have been found to be highly stable across grades (e.g., Cunningham & Stanovich, 1997; Nation & Snowling, 2004). Thus, predicting reading development and fostering precursors of reading proficiency are important for individual children as well as for modern societies as a whole. When considering preschool-age children's skills and abilities that may be most predictive of the development of individual differences in reading literacy, language competencies in particular have been found to be significantly associated with later attainment and success in reading literacy (e.g., Lervåg, Bråten, & Hulme, 2009; NICHD Early Child Care Research Network, 2005; Reese, Suggate, Long, & Schaughency, 2010). However, language competencies encompass a variety of – by no means homogeneous – abilities and skills. One distinction relevant to the prediction of later reading literacy is the differentiation between *phonological information processing skills* on the one hand and *linguistic abilities or more integrative language competencies*[1] on the other.

Reading-related *phonological processing skills* are often further subdivided into (a) phonological awareness skills (i.e., the sensitivity and ability to segment words into smaller units and to reflect on the structure of the sound of oral language), (b) phonological (working) memory (i.e., the individually different capacity to represent phonological information in working memory), and (c) fast access to phonological information in long-term memory, also known as rapid automatized naming (RAN) (see Torgesen, Wagner, & Rashotte, 1994).

Unlike phonological processing skills *linguistic abilities* refer more strongly to lexical knowledge (i.e., receptive and productive vocabulary and the structure of the lexicon) as well as to morphosyntactic knowledge (i.e., implicit knowledge of grammatical regularities). Closely related to these linguistic abilities or language components, more *integrative, functional language competencies* have to be considered. These require the mastery and integration of various language facets and come with higher ecological validity as they are closer to the everyday affordances of language that children are exposed to. Examples of such competencies are narrative discourse (e.g., the telling and retelling of a story), story comprehension, and sentence reproduction. They draw on various language facets including the child's lexical-semantic and grammatical knowledge as well as his or her phonological information processing skills.

However, although there is substantial research that has documented a close relation between language and reading, many questions concerning the specific and possibly different impacts of various language facets on reading literacy remain unresolved. This is presumably due to the complex multicomponential nature of language. In fact,

---

[1] In linguistics and the psychology of language, phonology is conceptualized as a subcomponent of linguistic knowledge (see Weinert & Grimm, 2008, 2012). However, in this article we use the term *linguistic abilities* to refer predominantly to vocabulary and grammar, thus differentiating *linguistic* knowledge from *phonological* processing skills. Furthermore, *integrative language competencies* refer to more comprehensive language measures that tap both, linguistic knowledge and phonological processing skills, and/or are closer to everyday language affordances (e.g. oral text comprehension, narrative discourse measures, sentence reproduction). Some authors use the term *oral language* (e.g., Muter, et al., 2004; Senechal, et al., 2006) to describe linguistic abilities in the above-mentioned sense as well as more integrative language competencies. However, because our aim is to differentiate between various aspects of oral language competencies (phonological processing, vocabulary and grammar, integrative language measures) we use the term *linguistic abilities* and *integrative language competencies* to refer to the respective aspect of oral language processing.

most studies have focused on only some of the potentially relevant facets of language in preschool-age children (e.g., vocabulary, phonological awareness) and have not taken into account the relative importance of these facets in the prediction of reading literacy. In addition, significant distinctions have to be made with regard to the outcome measure. Thus, the various aspects of language competencies are to be pitted against at least two different facets of reading literacy: the ability to decode written language (including measures of reading fluency) and the ability to comprehend (written) texts (Cain, 2010). When children begin learning to read, their initial task is to figure out how letters and written words map onto their phonological form. Thus, children have to discern the more or less regular grapheme-phoneme correspondence rules and to defragment them into phonological word forms. It is not until they have mastered this task that they can begin to read for meaning. Thus, children first have to master basic decoding processes before higher comprehension processes can take place. This holds true, and is even more pronounced, at the levels of sentences and texts.

Basic reading skills and improvements in decoding are often assessed by measures of reading accuracy. Whereas this is reasonable in orthographically inconsistent languages such as English, this is not the case in more consistent orthographies such as German. Here, a high level of reading accuracy is achieved very early in reading development, and the developmental progress in basic reading skills is better described as an improvement in fast and fluent reading as indicated by measures of reading fluency (Wimmer, 2006).

According to the simple view of reading, reading literacy is defined as a product of the processes of decoding and comprehension (Hoover & Gough, 1990). However, from a resource-allocation point of view, it is not only a person's decoding ability per se but, in particular, that person's ease and fluency of decoding (reading fluency) that seems to provide an important foundation for reading comprehension. Fluent readers probably need fewer resources for basic reading processes and thus they have more residual cognitive resources for processing and elaborating the information given in a text (e.g., Perfetti, 1985). In line with this assumption, reading fluency was shown to be a highly reliable predictor of reading comprehension (Kim, Wagner, & Foster, 2011).

Irrespective of this interrelation, the two facets of reading literacy (i.e., basic reading skills and reading comprehension) should be influenced differently by individual phonological processing skills and linguistic abilities or integrative language competencies. Acquiring basic reading skills affords the analysis and synthesis of strings of phonemes (i.e., phonological awareness), a comparison of the decoded phonemes with information stored in long-term memory (fast access to long-term memory), and the maintenance of the decoded phonemes in working memory. Thus, as far as basic reading skills are concerned, individual phonological information processing skills most likely play a major functional role. However, this might be different when reading comprehension is considered. In order to comprehend words, sentences, and texts, the reader has to draw on lexical knowledge (vocabulary), morphosyntactic knowledge (grammar), as well as text-specific formal and content-related knowledge. Obviously, when children begin to develop reading competencies, the written words, sentences, and texts presented to them tend to be very easy and thus might be understood with rather basic linguistic abilities; however, as decoding and reading fluency improve and children grow up, they begin to encounter and read more complex texts. At that time, advanced linguistic abilities should become more important for text comprehension. Yet, because reading comprehension affords a minimum of basic reading skills and is facilitated – via reduced cognitive load – by advanced basic reading skills, phonological processing skills may still have an (indirect) impact on reading comprehension.

In sum, when predicting reading literacy in school-age children from their language competencies in preschool, it is important to consider various language indicators as predictors; at the same time, different aspects of reading literacy should be taken into account as outcome criteria. However, studies differ in the language competencies that are assessed as well as in the reading outcomes measured in school-age children (e.g., decoding skills, reading fluency, reading accuracy, reading comprehension). Furthermore, these studies often refer to only some aspects of language and/or reading literacy. Thus, after a brief overview of empirical results regarding the predictive power of various facets of language for reading literacy in elementary-school-age children, we use data from the longitudinal BiKS-3-10 study to analyze the impact of various indicators of early language competencies on different facets of reading literacy in more detail.

Language Competencies as Predictors of Reading Literacy

The Role of Phonological Information Processing Skills in Learning to Read

A large amount of research has established the idea that phonological information processing skills are important predictors of individual differences in learning to read (e.g., Bryant, MacLean, Bradley, & Crosslan, 1990; Castles & Coltheart, 2004; Ehri, et al., 2001; Lonigan, et al., 2009; Muter, Hulme, Snowling, & Stevenson, 2004). To discern the more or less regular grapheme-phoneme rules of correspondence of the child's respective language, phonological awareness helps the child to find out how the sound structure of words maps onto the written words. Phonological awareness refers to "the ability to identify and manipulate the sound structure of words" (Cain, 2010, p. 76). Besides mapping the sound structure to written words, in order to read fluently, the child has to process phonological information in working memory and to quickly gain access to the phonological word forms stored in long-term memory in order to retrieve the respective word meaning. Thus, phonological information processing skills that are relevant for learning to read can be differentiated into phonological awareness, speed of access to verbal information in long-term memory, and phonological working memory capacity (Torgesen, et al., 1994; Wagner & Torgesen, 1987; see also Cain, 2010). However, although these facets are related, they are not identical and may have different impacts on reading development.

Phonological awareness. Phonological awareness comprises the sensitivity and ability to reflect on and be aware of the sound structure of language. To assess phonological awareness, children are usually asked to delete, count, or substitute sound units (analysis tasks), to combine sounds (synthesis tasks), to match sounds within words (identity tasks), or to respond to rhyming tasks (produce a word that rhymes or judge whether pairs of words or nonwords rhyme or not; Shanahan & Lonigan, 2010). Intervention studies in different countries have shown that children trained in phonological awareness skills such as rhyming or segmenting words into phonemes outperform untrained children on measures of phonological awareness as well as in later reading and writing (e.g., Bradley & Bryant, 1985; Lundberg, Frost, & Peterson, 1988). Training effects have also been found for at-risk children who show poor phonological awareness or language skills (e.g., Berendes, 2011; Schneider, Ennemoser, Roth, & Küspert, 1999), have immigration backgrounds (Souvignier, Duzy, Glück, Pröscholdt, & Schneider, 2012), or come from families with low socioeconomic status (Ehri, et al., 2001; Lundberg, Larsman, & Strid, 2012).

Some researchers argue that phonological awareness is less important in languages with regular or more consistent orthographies compared to orthographically less consistent languages. In line with this argument, training programs as well as longitudinal studies conducted in countries with a regular orthography such as Finland, Germany, the Netherlands, or Norway have shown that phonological awareness has an effect on the early stages of reading in particular (e.g., de Jong & van der Leij, 1999; Lervåg, et al., 2009), whereas in English-speaking countries, effects have been demonstrated for longer periods in reading development (e.g., Byrne & Fielding-Barnsley, 1995; Muter, et al., 2004; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). These results suggest that phonological awareness is more closely related to basic reading skills than to reading comprehension.

Fast access to phonological information in long-term memory. The ability to quickly access phonological information stored in long-term memory is thought to facilitate reading because the child has to match written words with sounds stored in long-term memory. Indeed, children who exhibit poor reading skills often show deficits in the ability to access phonological information in long-term memory (Morris, et al., 1998; Stanovich & Siegel, 1994). In the same vein, individual differences in the ability to quickly access phonological information are correlated with differences in reading acquisition (Torgesen, et al., 1999).

To assess how rapidly children are able to access phonological information in long-term memory, they are usually asked to name well-known objects, letters, or digits as quickly as possible (RAN (rapid automatized naming) tasks). These rapid-naming measures have been shown to impact early reading literacy even when other measures of phonological processing are statistically controlled (Kirby, Parrila, & Pfeiffer, 2003; Lervåg, et al., 2009; Schatschneider, et al., 2004). Thus, there is empirical evidence indicating that rapid automatized naming is a reliable predictor of reading literacy. In particular, in more orthographically consistent languages such as German or Greek, rapid automatized naming seems to be more important for basic reading processes than phonological awareness (Georgiou, Rauno, & Papadopoulus, 2008; Wimmer, Mayringer, & Landerl, 1998).

Phonological working memory. Besides phonological awareness and the ability to quickly access long-term memory, phonological working memory has been discussed as being important for reading development (e.g., Lonigan, et al., 2009; Wagner, et al., 1997). The capacity of phonological working memory is usually assessed through digit or word-span tasks or by using nonword repetition tasks. In these tasks, the child has to immediately repeat orally presented material that differs in length and/or complexity.

Torgesen et al. (1994) among others have reported medium to high correlations between phonological working memory performance in preschool and later reading ability. In the same vein, Ennemoser, Marx, Weber, and Schneider (2012) found almost identical correlations between measures of phonological working memory and various facets of reading literacy on the one hand and between phonological awareness tasks and these reading outcomes on the other. However, because the various aspects of phonological processing seem to share a large amount of common variance, measures of phonological working memory do not seem to account for unique variance in basic reading skills as assessed by measures of reading accuracy when other indicators of phonological processing are statistically controlled (Lervåg, et al., 2009; Torgesen, et al., 1994).

Phonological working memory may also have an *indirect* effect on later reading literacy mediated through linguistic abilities. In fact, it has been documented that early lexical learning is significantly influenced by phonological working memory capacity (e.g., Ebert, et al., 2013; Gathercole & Baddley, 1989; Weinert, Ebert, Lockl, & Kuger, 2012). Because lexical learning is expected to be important for later measures of reading development, studies that focus on the early stages of reading instruction may miss this effect.

In sum, it is well documented that phonological processing skills are significantly, although partially redundantly, associated with learning to read. However, the impact of the various indicators of phonological information processing seems to vary – at least partially – according to the orthographic consistency or inconsistency of the language (Georgiou, et al., 2008). For instance, in a study comparing German- and English-speaking children, Mann and Wimmer (2002, cited in Georgiou, et al., 2008) showed that phonological awareness was the only significant predictor of reading fluency in English-speaking children, whereas for German-speaking children, only RAN measures turned out to be predictive Georgiou et al. (2008) demonstrated that phonological awareness was a better predictor of decoding skills in English than in Greek children.

In theory, phonological processing skills should be associated with decoding processes in particular (see Cain, 2010). Therefore, studies that have investigated the impact of phonological processing skills have predominantly focused on decoding and basic reading skills such as reading accuracy and reading fluency or reading speed. At the same time, phonological processing skills are interconnected with linguistic abilities and thus may have an additional indirect effect on later reading literacy, especially on reading comprehension. As argued in more detail in the next section, linguistic abilities are also correlated with reading literacy and with reading comprehension in particular.

The Role of Linguistic Abilities and More Integrative Language Competencies in Reading Development

Linguistic abilities and more integrative language competencies are important for later reading literacy for various reasons (e.g., Dickinson, McCabe, Anastasopoulus, Peisner-Feinberg, & Poe, 2003; Muter, et al., 2004; Reese, et al., 2010). Obviously, (written) text comprehension presupposes lexical and semantic as well as morphosyntactic knowledge to enable a person to understand (written) words and sentences and the meaning of texts. Grammatical and semantic knowledge also help a person to unravel unknown words and to infer the exact interrelations between propositions. The more complex a written text is (e.g., including challenging vocabulary and sentence structures), the more linguistic knowledge is required to decipher its meaning. Besides vocabulary and grammatical knowledge (i.e., language components), more integrative and functional language competencies may be of special relevance to (written) text comprehension. In particular, children's narrative skills and comprehension skills for oral texts may play important roles. However, most studies that have predicted reading literacy not only through phonological processing skills have focused on vocabulary or syntactic abilities, whereas only a few have examined functional, more integrative language competencies such as oral text and discourse comprehension (listening comprehension of orally presented texts/discourses) (Cain, 2010).

Studies that have taken *vocabulary* into account have often demonstrated that it has a significant relation to later reading literacy (e.g., de Jong & Leseman, 2001; de Jong & van der Leij, 2002; Muter, et al., 2004; NICHD Early Child Care Research Network, 2005). Correlations have been found between vocabulary and early basic reading skills (e.g., NICHD Early Child Care Research Network, 2005) as well as early reading comprehension (e.g., Roth, Speece, & Cooper, 2002). The strongest effects, however, have been documented between vocabulary and later reading – specifically for later reading comprehension (e.g., de Jong & van der Leij, 2002; Sénéchal & LeFevre, 2002). De Jong and van der Leij (2002), for example, documented an effect of individual differences in vocabulary on later reading comprehension at the age of 10, even when controlling for reading comprehension at the age of 7. Thus, their study was able to demonstrate that vocabulary is correlated not only with later reading comprehension, but has an effect on its growth as well.

Interestingly, Ouellette (2006) showed – based on a study of 60 children from the fourth grade – that receptive vocabulary (breadth of vocabulary) is specifically relevant for decoding, whereas depth of lexical knowledge (select synonyms, providing definitions) impacts reading comprehension. Similarly, Roth et al. (2002) reported comparatively higher correlations between tasks requiring oral word definitions and reading comprehension compared to those between receptive vocabulary and reading comprehension. These results suggest that vocabulary (and specific aspects of lexical knowledge) may have differential and various effects on reading literacy.

Some studies have considered not only vocabulary but also additional linguistic abilities. However, they frequently distinguished only between vocabulary and a broad language measure, which comprises different language measures such as oral text comprehension and expressive language skills. For example, in a longitudinal study following more than 600 children from preschool to grade 4, Storch and Whitehurst (2002) found a direct effect of receptive vocabulary on concurrent reading comprehension (Grade 3 & 4) and an indirect effect of a broad language measure (including oral text comprehension) assessed in preschool on word reading measured in Grade 1 and Grade 2 mediated by code-related skills (phonological awareness, letter knowledge). Their study thus demonstrated direct and indirect effects of linguistic measures on reading literacy. However, their study did not address which of the various language aspects was most important for reading literacy.

Direct and indirect influences of linguistic abilities or broad language measures on early reading literacy were also demonstrated by another comprehensive study conducted by the NICHD Early Child Care Research Network (2005). Interestingly enough, early vocabulary and a broad language measure (including various language indicators such as measures for syntax, oral text comprehension, and expressive language) were found to covary with decoding to nearly the same extent in the first grade; but when both measures were included in the model, only the broad language measure predicted decoding in the first grade and thus played the more prominent role. Specifically, the study found that the broad language measure at the age of 4 was directly associated with decoding skills in the first grade. Furthermore, the study identified significant indirect paths from the broad language measure at the age of 3 as well as from the age of 4. In contrast to the prediction of first graders' decoding skills, when predicting reading comprehension, not only did the earlier broad language measure provide a direct path, but also vocabulary. However, reading comprehension was measured in the third grade and more basic reading skills were measured in the first grade. Thus, it was not possible to judge the effect of the various oral language measures on reading comprehension compared to more basic reading skills at the same developmental time point.

Besides broad language measures, which comprise various indicators of receptive and productive language facets and competencies, some studies have focused more specifically on functional or integrative language measures (i.e., narrative skills or oral text comprehension) that are ecologically valid and/or conceptually connected to reading comprehension. However, results concerning the impact of these more integrative measures on reading literacy are heterogeneous. For example, a study by de

Jong and Leseman (2001) revealed that vocabulary and listening comprehension (i.e., oral text comprehension) in Grade 1, when considered separately, were significant predictors of reading comprehension in Grade 3 when controlling for early reading comprehension and word decoding as well as for home literacy and problem solving. However, when accounting for both language indicators simultaneously, only vocabulary had a significant effect. By contrast, de Jong and van der Leij (2002) did not find that vocabulary in Grade 1 accounted for additional variance in reading comprehension in Grade 3 when controlling for listening comprehension (i.e., oral text comprehension) and reading comprehension in Grade 1. This may be due to the fact that more integrative language skills such as oral text comprehension become more and more important as reading skills continue to develop. Reese et al. (2010), for example, found that after 1 year of reading instruction, children's narrative skills did not predict their concurrent reading skills (i.e., reading fluency) when differences in early decoding skills were accounted for. In a second study, however, they showed that after 2 years of reading instruction, the quality of children's narratives predicted their concurrent reading skills as well as reading skills 1 year later, even after controlling for vocabulary and early decoding skills.

In addition to measures of narrative discourse and oral text comprehension, which are accepted as functional and conceptually relevant to reading comprehension, another integrative language measure (i.e., sentence reproduction) has been found to be highly predictive of reading literacy as well. In a German longitudinal study that included 53 children, sentence reproduction in the last year of preschool attendance (i.e., at age 5) was found to be the best predictor of children's basic reading skills in the second year of formal reading instruction compared to other language measures including phonological awareness (Goldammer, Mähler, Bockmann, & Hasselhorn, 2010). However, the theoretical status of sentence reproduction tasks is controversial. Sometimes these tasks are classified as memory tasks, sometimes as integrative measures of vocabulary and phonological processing, and sometimes as indicators of grammatical knowledge (especially when vocabulary is rather easy and the sentence exceeds memory span, which is normally the case in these tasks; see Weinert, 2010b). The ambiguous theoretical status is due to the fact that sentence reproduction tasks tap various language skills. Although sentence reproduction tasks are less ecologically valid, they draw from available lexical-semantic and grammatical knowledge that help the child to process, represent, and recall/reconstruct the content and structure of a given sentence and to hold it in short-term memory even when the number of words exceeds capacity restrictions. At the same time, because working memory is involved in this task to a large degree, the task also taps phonological processing skills. This may explain the strong impact of sentence reproduction on basic reading skills. However, the impact on reading comprehension remains an open question. Furthermore, it is not clear whether this kind of measure is a better predictor of reading literacy than more ecologically valid measures such as oral text comprehension.

Comparing Phonological Information Processing Skills and Linguistic Abilities or Integrative Language Competencies as Predictors of Reading Literacy

The studies summarized so far demonstrate that both phonological information processing skills and linguistic abilities in the sense of vocabulary and/or grammar but also more integrative language measures are reliable predictors of later reading literacy. Thus, the question arises whether phonological information processing or linguistic as well as more integrative language measures are more important to the development of reading literacy.

An extensive meta-analysis of about 300 published articles carried out by the National Early Literacy Panel (NELP; see Shanahan & Lonigan, 2010) indicated that specifically phonological awareness and rapid automatized naming of letters/digits/objects/colors showed moderate to large effects in predicting later basic reading skills (decoding) and reading comprehension. Somewhat smaller but still moderate were the effects of phonological memory in predicting basic reading skills (decoding) and reading comprehension. The covariation of reading literacy and phonological processing variables was maintained when differences in other variables, such as IQ or socioeconomic status, were accounted for. By contrast, the ability to produce and comprehend oral language did not always preserve its predictive power when other variables were controlled, although this ability was also moderately to highly correlated with later basic reading skills (i.e., decoding) and reading comprehension. Furthermore, the results suggest that linguistic abilities are more important when more complex or broad measures instead of simple vocabulary measures are considered. Moreover, complex integrative or broad language measures were more strongly associated with reading comprehension (about $r = .70$) than with basic reading skills (i.e., decoding; about $r = .58$). For vocabulary measures, this difference in predictive power was not observed (Shanahan & Lonigan, 2010). These are important results as studies often assess only vocabulary and therefore may underestimate the effect of linguistic abilities and more integrative language measures (see also Dickinson, Golinkoff, & Hirsh-Pasek, 2010).

In sum, the meta-analysis suggested that there are effects of phonological information processing skills as well as of linguistic abilities and more integrative language measures on later reading achievement, although the correlations between phonological information processing skills and reading literacy seemed to be more robust across studies and less affected by methodological variations. However, Dickinson et al. (2010) criticized this NELP report as failing to adequately recognize the role of linguistic abilities and more integrative language competencies. They argued that the meta-analysis failed to consider indirect effects of these measures on later reading literacy. For example, Sénéchal, Ouellette, and Rodney (2006) demonstrated an effect of vocabulary on gains in phonological awareness, which was found to be one of the strongest predictors in the above-cited meta-analysis. Furthermore, Dickinson et al. (2010) argued that linguistic abilities and more integrative language competencies, in contrast to phonological abilities, develop over an extended period of time and therefore have longer lasting effects that were not considered in the time period included in the meta-analysis. In this vein, a Finnish study revealed the strongest (indirect) predictive links between linguistic abilities in preschool and reading fluency and accuracy at 9 years of age for receptive and expressive language via measures of letter naming, morphology, and phonological awareness. However, direct links were stronger for phonological information processing skills such as rapid naming and phonological sensitivity (Torppa, Lyytinen, Erskine, Eklund, & Lyytinen, 2010). Moreover, most studies that have considered both phonological information processing skills and linguistic abilities have shown that phonological awareness had a stronger effect on early reading literacy, whereas linguistic abilities had more impact on later reading literacy, especially reading comprehension (e.g., NICHD Early Child

Care Research Network, 2005; Sénéchal & LeFevre, 2002; Sénéchal, et al., 2006; Wagner, et al., 1997). A recent German study enhanced these conclusions by comparing the results of two German longitudinal studies that both included measures of reading fluency and reading comprehension. Both studies showed that linguistic abilities were more strongly connected to later reading and, in particular, to reading comprehension, whereas phonological processing turned out to be more strongly connected to early reading achievement (reading fluency as well as reading comprehension; Ennemoser, et al., 2012). Moreover, Sénéchal et al. (2006) demonstrated that vocabulary and oral text comprehension explained a unique proportion of variance in reading comprehension in Grade 3 but not in Grade 1 (when accounting for parents' education, earlier reading comprehension, early literacy, and phonological awareness in kindergarten). On the other hand, phonological awareness was found to be a stronger predictor of reading comprehension in Grade 1 than in Grade 3. In a second study, Sénéchal et al. (2006) showed similar results for French-speaking children. Results revealed that receptive vocabulary measured in kindergarten had an effect on reading comprehension in Grade 4 after accounting for various variables such as word reading in Grade 1 and reading fluency in Grade 4, but not on reading fluency after accounting for reading comprehension, parents' education and literacy, early literacy, and phonological awareness.

In sum, the literature suggests that phonological information processing skills are especially important for early reading development, particularly when basic reading skills such as decoding and reading fluency are concerned; linguistic abilities and more integrative language competencies, however, seem to play a major role in later reading development, particularly in reading comprehension. Although this seems to be a straightforward suggestion when considering models of learning to read, the issue is actually more complicated because linguistic abilities themselves build upon phonological information processing and vice versa. Specifically, early lexical learning and vocabulary acquisition draw heavily on phonological knowledge as well as on phonological working memory capacity, i.e., phonological working memory is an important predictor of early vocabulary growth (Weinert, 2010a; see also Ebert, et al., 2013; Weinert, et al., 2012). However, from the age of 6 onwards (or even earlier), vocabulary has been shown to be predictive of the growth of phonological working memory (Gathercole, et al., 1992). Furthermore, integrative language measures (e.g., oral text comprehension) draw on linguistic abilities such as vocabulary and grammar as well as on phonological processing skills. In addition, the phonological sensitivity approach states that vocabulary provides the foundation for phonological sensitivity and awareness, which in turn support early reading development and decoding skills (see Dickinson, et al., 2003; Sénéchal, et al., 2006). Thus, the various language skills and measures seem to be highly interconnected in the preschool years and appear to influence each other. Accordingly, Dickinson et al. (2003) foster a comprehensive language approach suggesting that various language abilities and skills, such as phonological information processing and linguistic abilities including integrative oral language competencies, are interrelated during the preschool years and that these relationships persist in later reading development.

Taken together, phonological information processing skills and specifically phonological awareness (at least in orthographically more inconsistent languages such as English) seem to have a comparatively strong impact on reading literacy. By contrast, the influence of linguistic abilities and integrative language competencies is more diversified. These become more strongly related to reading literacy during the course of reading acquisition in the early school years and their effects are not only direct but also indirect through phonological information processing and thus probably through basic reading skills as well. Furthermore, linguistic abilities (vocabulary, grammar) and more integrative language competencies seem especially important for reading comprehension, whereas phonological processing skills are more important for basic reading processes such as decoding or reading fluency. However, the results are not totally clear. Some studies have also revealed that linguistic abilities and more integrative language measures are correlated with basic reading skills, whereas phonological processing skills are correlated with reading comprehension. As outlined, an explanation for these findings might be that phonological processing and linguistic abilities are strongly interconnected. Phonological working memory, for example, is predictive of early vocabulary development, whereas later, vocabulary is itself predictive of the growth of phonological working memory (Gathercole, et al., 1992) and phonological awareness (Sénéchal, et al., 2006). Furthermore, integrative language measures tap not only linguistic abilities but also phonological information processing skills. Thus, the question is whether phonological processing skills and linguistic abilities are separable at all in the early years or whether they represent a single construct of global language competencies in general.

In summary, various studies have stressed the importance of language competencies in the development of reading literacy. Some researchers have more strongly referred to phonological information processing as an important predictor of later reading literacy, whereas others have emphasized linguistic abilities (vocabulary, grammar), more integrative language measures, or broad language measures (summing across various indicators and facets). The literature suggests that both phonological information processing skills and linguistic abilities or more integrative language measures are of relevance to reading development but seem to influence reading literacy in different ways and at different time points in development. Phonological processing has been found to be more relevant to basic reading skills such as decoding and reading fluency and in early phases of reading development, whereas linguistic abilities and integrative language measures have demonstrated a stronger impact on reading comprehension and on later reading development.

However, studies differ in the language competencies that are assessed as well as in the reading outcomes measured in school-age children (e.g., decoding skills, reading fluency, reading accuracy, reading comprehension). Furthermore, these studies often refer to only some aspects of language and/or reading literacy. Thus, empirical results concerning the impact of various language skills for reading literacy are heterogeneous and ambiguous. The present study considers phonological processing skills and linguistic abilities in early preschool-age children and tests for their predictive effects on (a) more basic reading skills (reading fluency) and (b) reading comprehension. Furthermore, although much is known about the impact of phonological processing skills on reading literacy and on basic reading skills in particular, less is known about the relative impact of lexical, grammatical, and/or more functional and integrative language competencies on more advanced reading competencies such as reading comprehension. This may be due to the fact that only a few studies to date have considered and systematically differentiated various linguistic abilities and language measures. Thus, the present study addresses this issue in depth by analyzing the contributions of lexical-semantic, grammatical, and more integrative, functional language measures on reading comprehension. Moreover, most studies have assessed these early predictors of reading in the last year before school entrance. Thus, we know little about the impact of early language skills on later reading literacy, but this link is especially important to uncover because phonological processing skills and linguistic abilities are strongly interrelated and influence each other over the course of development.

Therefore, we (1) analyzed whether phonological processing skills and linguistic abilities could be separated in early preschool-age children and – if so – (2) tried to replicate the finding that linguistic abilities are especially relevant to reading comprehension, whereas phonological processing skills are more predictive of basic reading skills. In this vein, we investigated whether this would even be true when language competencies were assessed early in the preschool years and for early reading comprehension in Grade 2 when reading literacy is just beginning.

Because less is known about the relative impact of various indicators of linguistic abilities and more integrative language competencies for reading literacy, we (3) further focused on reading comprehension and its prediction through various linguistic abilities and integrative language measures. (a) First, we asked which linguistic component – vocabulary (assessed in most studies) or grammar (often not assessed as a separable linguistic component) – would have a comparatively stronger impact on early reading comprehension. (b) Additionally, we investigated whether integrative and functional measures of early language competencies would explain additional variance over and above linguistic abilities in the sense of language components such as vocabulary and grammar.

Method

Procedure and Sample

Data were drawn from the German BiKS-3-10 study (see for more information about BiKS-3-10 Lorenz, Schmitt, Lehrl, Mudiappa, & Rossbach, chapter 2, this volume). The sample in the present study was comprised of children who had been participating in the BiKS-3-10 study since they were about 3 years old (N = 554).[2] At this age, most of the children had just started preschool. In this study, we focused on children's language competencies in the first and second year of preschool (measurement points 2 and 3 of the BiKS-3-10 study) as well as on their reading literacy in Grade 2 (about 3 years later). At measurement point 3, when most of the language tests relevant for this study were administered, children were about 4;8 years old (SD = 4.47 months). Their families' highest international socioeconomic status (HISEI; see Ganzeboom, de Graaf, & Treiman, 1992, for further information) was on average 52.2 (SD = 16.3). With regard to parents' mother tongue, 12.1% of the children had parents who both spoke a different first language than the lingua franca of society (German), whereas 9.7% lived in families with one parent who had a mother tongue other than German.

Preschool-age children were tested individually in separate rooms at their preschools. After entry into the formal school system, testing took place in small groups in school or individually at home depending on the measure assessed. All assessments were conducted by extensively trained students using – as much as possible – standardized tests with approved quality.

Measures

For preschool-age children, various language measures were assessed. At measurement point 3 of the BiKS-3-10 study (age: 4;8 years), children completed two tests measuring phonological processing skills (phonological working memory; rapid naming) and two tests assessing linguistic competencies (receptive vocabulary; receptive grammar). A subgroup of 128 children[3] received two additional tests measuring integrative (functional) language competencies (reproduction and comprehension of an orally presented story; sentence reproduction). Sentence

---

[2] Seven of these children entered the study at a later time point because they started preschool after our first measurement point, but like the other children in our study, they were expected to enter school in autumn 2008.

[3] At measurement point 3 of the BiKS-3-10 study, this subgroup of children was 4;9 years old (M = 57.02 months, SD = 2.06). About 7.0% of these children had parents who both spoke a mother tongue other than German, and about 4.7% had one parent with a mother tongue other than German. The mean HISEI of this subsample was 52.3 (SD = 14.9).

reproduction was assessed at measurement point 3, whereas story reproduction and comprehension were assessed at measurement point 2, about half a year earlier.

Phonological processing skills

Phonological working memory. Children completed a digit span task taken from the German Version of the Kaufman Assessment Battery for Children (K-ABC; Melchers & Preuss, 2003). Children had to reproduce sequences of digits ordered in sets of increasing length. Each set consists of three items made up of the same number of digits. Testing ends when children fail to correctly reproduce a single item in a set. For each correctly recalled item, children receive 1 point. The number of correctly recalled items was used in the analyses.

Rapid naming. To assess children's fast access to phonological information stored in long-term memory, a rapid naming task was administered. Children had to name five familiar objects: Eis (ice), Ball (ball), Hund (dog), Baum (tree), Fisch (fish) as fast as possible. These objects were presented on a picture card and the pictures were repeatedly presented in a random order in five rows. The time the child needed to name all objects on the sheet was used for the analyses.

Linguistic measures

Vocabulary. Receptive vocabulary was assessed by an unpublished German Research Version of the Peabody Picture Vocabulary Test - Revised (PPVT-R; Dunn & Dunn, 1981; Research Version: Roßbach, Tietze, & Weinert, 2005). Children were presented individual words accompanied by four black-and-white pictures per item. The test consists of 175 items clustered in sets of 12 items (last set 7 items). The children's task is to point to the picture that depicts the meaning of the orally presented word. Testing ends when children answer six or more items per set incorrectly. The total number of correct items was used in the analyses.

Grammar. To assess children's receptive grammar, a short version of the German Version of the Test for the Reception of Grammar (TROG; Bishop, 1983/1989; German Version: TROG-D; Fox, 2006) was implemented. Children are orally presented with sentences accompanied by four colored pictures per sentence. Their task is to select the picture that corresponds to the stimulus sentence. Items are grouped in sets.

The first three sets control for vocabulary. The 18 sets that follow are comprised of sentences of increasing grammatical complexity with two items per sentence structure. Testing ends when children answer five succeeding sets incorrectly; a set is counted as failed when at least one item is answered incorrectly. Each correct answer was scored as 1 point, and a maximum of 48 points could be received.

Integrative language measures

Story reproduction and comprehension. To assess children's story reproduction and comprehension, we used a version of a Scottish fairy tale employed in a number of psychological studies (e.g., Wimmer, 1982). In this fairy tale, a farmer wants to bring his donkey into the barn, but the donkey doesn't want to go. So the farmer asks his dog to bark so that the donkey will get frightened and run into the barn. The story ends with the dog barking and the donkey running into the barn. After a short delay, children were asked to reproduce the story. For motivation, a teddy bear was introduced to listen to the child's reproduction. Subjects were prompted to tell as much about the story as they could remember. If they did not begin to retell the story, up to three general prompts were provided (e.g., "What happened in the story?"). If children stopped during their retelling of the story, again, general prompts were given (e.g., "Tell me more"; "What happened then?"). As a first measure of the children's story reproduction, the number of propositions (content units) recalled was counted. Children could receive up to 11 points. After finishing their free recall, children were asked specific questions about the story. These questions consisted of three "What questions" and three "Why questions" (e.g., "What should the dog do?"; "Why did the farmer want the dog to bark?"). Each correct answer was scored as 1 point. Thus, children could receive a maximum of 6 points.

Sentence reproduction. As another integrative measure of early oral language competencies that draws on lexical and grammatical knowledge as well as on phonological processing skills, the subtest "Sentence Memory" of a German language battery for children (SETK 3-5: Sprachentwicklungstest für drei- bis fünfjährige Kinder; Grimm, 2001) was administered. In this task, the children were presented with 15 sentences of increasing grammatical complexity and length, and they were asked to immediately reproduce each sentence. About half of the sentences were semantically incongruent (i.e., nonsense sentences, e.g., “The stupid parrot knits on the bottle”). Thus, some of the sentences drew on linguistic knowledge as well as on world knowledge whereas others drew specifically on linguistic knowledge (grammar, vocabulary). Each sentence reproduction was scored according to the number of words correctly recalled. In total, the children could receive 119 points.

Reading literacy

All children who still took part in the BiKS-3-10 study in Grade 2 of primary school were administered two tests of reading literacy, one of them assessing basic reading skills (reading fluency/speed) and the other reading comprehension.

Basic reading skills (reading fluency/speed). As a measure of the children’s basic reading skills, the SLS 1-4 (Salzburger Lese-Screening für die Klassenstufen 1-4; Mayringer & Wimmer, 2003) was administered. Children are instructed to read as quickly as possible a series of simple sentences with increasing length. The child has to evaluate whether the content of the sentence he or she just read is true or false. Because each statement (sentence) is very obviously true or false, the evaluation of its truth should be easy (e.g., “Bananas are blue”). The number of sentences judged correctly within 3 min is assessed. According to the authors, this test measures basic reading skills in a natural reading context with a focus on reading speed.

Reading comprehension. For assessing reading comprehension, the subtest “text comprehension” of a German reading literacy test for first to sixth graders (ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässler; Lenhard & Schneider, 2006) was implemented. Children had to read short passages and to answer one to three multiple-choice questions about each passage. Each multiple-choice question provided four alternative answers. The questions tapped either information given explicitly in the text or they required the child to extract meaning or to draw inferences from the text. Children received 1 point for each correctly answered multiple-choice question with a maximum of 20 points.

Statistical Analyses

Subsamples considered in the analyses. When focusing on reading literacy, children who were enrolled in school at time points that differed from the main sample (N = 54) had to be excluded from the analyses because of different levels of formal reading instruction. Furthermore, not all children of the cohort sampled in preschool could be followed until they were school age. Thus, only those children who were tested for reading literacy in Grade 2 ($N = 293$) were included in these analyses.

According to the study design, language measures testing for integrative, functional language competencies were assessed only in a subgroup of 128 children. Thus, analyses of these measures refer to this subgroup of children. Again, children were excluded from analyses concerning reading literacy in school if they were enrolled in school at time points that differed from the main sample ($N = 7$), and only those children who were tested for reading literacy in Grade 2 were included in the analyses ($N = 74$).

Procedure. In the following, we first refer to descriptive statistics for the two subsamples before evaluating two alternative models (a one- and a two-factor model) of children's language competencies in preschool using confirmatory factor analyses. Based on these results, reading literacy was predicted by children's language competencies. For these analyses, the full-information-maximum-likelihood (FIML) approach (e.g., Arbuckle, 1996) implemented in Mplus Version 6.0 (Muthén & Muthén, 2010) was adopted to deal with missing data. This approach includes valid information of all observations to estimate model parameters.

In a second step, more specific analyses were conducted to determine the relative impact of vocabulary and grammar when predicting reading literacy by using hierarchical regression analyses. The uniquely explained variance was estimated by entering the corresponding variable (vocabulary or grammar, respectively) in the last step to test for the specific proportion of variance explained by these predictors.

Finally, in a third step, we focused on the role of integrative, functional measures of early (oral) language competencies and their abilities to predict reading literacy after controlling for vocabulary and grammar. Again, hierarchical regression analyses were used to test for the specific contribution of these language measures to later reading literacy.

Results

Descriptive Statistics

Table 1 shows the descriptive statistics for language measures in preschool and for reading literacy in Grade 2 relevant for the present study. Statistics are presented separately for the whole sample and the subgroup of children who were given additional tests on integrative language competencies.

Table 1. Descriptive Statistics for the Total Sample and a Subgroup of Children who were Additionally Tested on their Integrative and Functional Language Competencies

	Total Sample			Subgroup		
Measures	*N*	*M*	*SD*	*N*	*M*	*SD*
Age, time 3 (preschool)	519	55.7	4.5	117	57.1	2.1
Age, Grade 2	298	97.8	4.0	78	100.0	2.3
Phonological Processing Skills						
Phonological Working Memory						
Digit Span (ZN, K-ABC), time 3	519	5.7	2.3	117	6.3	2.0
Access to long-term memory						
Rapid Naming, time 3	495	32.3	10.0	111	30.2	9.0
Linguistic Abilities						
Vocabulary						
PPVT, time 3	504	56.0	21.7	114	59.7	18.8
Grammar						
TROG, time 3	518	30.6	7.1	117	32.1	6.4
Integrative Language Competencies						
Sentence Reproduction						
Sentence Memory (SETK 3-5), time 3				106	80.3	20.6
Story Reproduction & Comprehension						
Story Reproduction, time 2				123	2.2	2.7
Story Comprehension, time 2				122	3.8	1.9
Reading Literacy						
Reading Speed (SLS 1-4), Grade 2	296	31.6	10.0	76	33.1	10.4
Reading Comprehension (ELFE 1-6), Grade 2	248	10.0	4.3	64	11.1	4.6

Note. ZN = Zahlennachsprechen (digit span); K-ABC = Kaufman Assessment Battery for Children; PPVT = Peabody Picture Vocabulary Test; TROG = Test for the Reception of Grammar; SETK 3-5 = Sprachentwicklungstest für 3-5jährige Kinder (language test battery); SLS 1-4 = Salzburger Lesescreening für die Klassenstufen 1-4 (reading speed); ELFE 1-6 = Ein Leseverständnistest für Erst- bis Sechstklässler (reading comprehension).

Due to the study design, this subsample was more homogenous in age and was on average 1.5 - 2 months older. Therefore, these children scored somewhat higher on all language measures than the total sample. For both samples, the number of children varied with respect to the measures assessed. This was mainly due to absences on the day of testing because testing took place on up to 4 days per measurement point. Concerning reading literacy in Grade 2, differences in sample sizes were due to the fact that ELFE (reading comprehension) was assessed in school, whereas SLS (reading fluency/speed) was tested at home. Although some schools refused to take part in the study, we were able to test children at home. Despite rather high stability in the sample, some children were lost because their families removed, they ended up attending special schools (e.g., Waldorf), their families lost interest in taking part in the longitudinal BiKS study, or for other reasons. However, in Grade 2, there were still 326 children who were tested for basic reading skills (SLS 1-4) and 263 children for reading comprehension (ELFE 1-6).

Table 2 shows moderate to high correlations between phonological processing measures, linguistic measures, and reading literacy for the whole sample. As predicted, all language measures were significantly correlated, although their covariations with rapid naming were only moderate. The intercorrelations between digit span as an indicator of phonological memory and the other measures were somewhat higher, whereas those between vocabulary or grammar and the others were quite similar. The highest correlation was found between the linguistic variables (i.e., vocabulary and grammar).

Table 2. Correlations between Measures of Phonological Processing, Linguistic Abilities, and Reading Comprehension for the Total Sample

	1	2	3	4	5
1. Digit Span, time 3					
2. Rapid Naming, time 3	-.26**				
3. Vocabulary, time 3	.43**	-.26**			
4. Grammar, time 3	.45**	-.22**	.63**		
5. Reading Comprehension, Grade 2	.30**	-.20**	.34**	.26**	
6. Reading Speed, Grade 2	.29**	-.33**	.18**	.18**	.80**

Note. Correlations between rapid naming and the other measures are negative because the score on the measure is the time needed to complete the task.
** $p < .01$

Furthermore, Table 2 shows that basic reading skills (reading speed) and reading comprehension were highly interrelated (r = .80). Measures of grammar and vocabulary were more strongly associated with reading comprehension than with basic reading skills, whereas phonological processing skills (digit span, rapid naming) were correlated with both basic reading skills and reading comprehension, although the correlations with basic reading skills were slightly higher.

Focus 1: Early Phonological Processing Skills and Linguistic Abilities

Are early phonological processing skills and linguistic abilities two distinguishable facets of language in preschool? Concerning our first research question (i.e., the separability of phonological processing skills and linguistic abilities in early preschool-age children), confirmatory factor analyses (CFA) were conducted. Two alternative models, a one-factor and a two-factor model, were evaluated and compared. Thus, we analyzed whether it would be statistically possible to differentiate phonological processing skills and linguistic abilities as two distinct though correlated dimensions of language processing at the age of 4 years or whether these facets are better described as indicators of one global dimension of language competence. The one-factor model combined all language measures, that is, vocabulary (PPVT), grammar (TROG), rapid naming, and phonological memory (digit span), as indicators of one global factor. The two-factor model consisted of two different factors, one for *Phonological Processing* and one for *Linguistic Abilities*. The factor *Phonological Processing* was indicated by the measure of phonological memory (digit span) and by the measure of the ability to quickly access phonological representations in long-term memory (rapid naming). The factor *Linguistic Abilities* was indicated by children's vocabulary (PPVT) and grammar (TROG). Models were evaluated using the statistical software Mplus version 6.0 (Muthén & Muthén, 2010). The full-information-maximum-likelihood (FIML) approach implemented in Mplus was used to adjust for missing data.

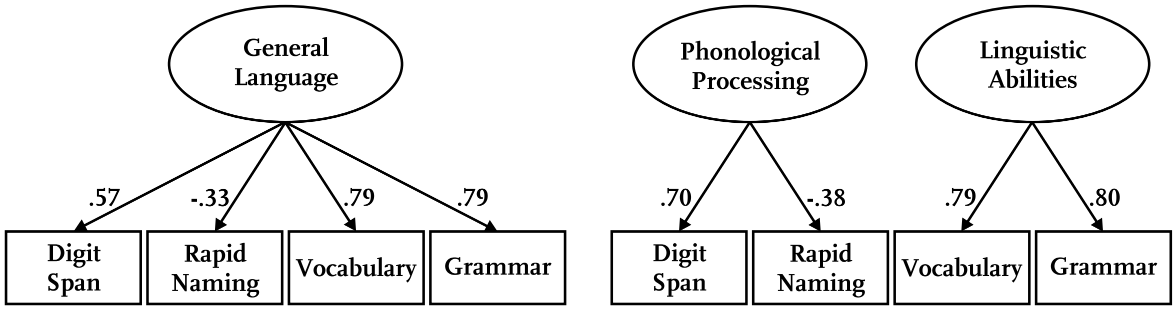

Figure 1. One-factor and two-factor models depicting factor loadings and correlations between factors at measurement point 3 (4;8 years). Circles represent latent variables and rectangles represent observed variables. All values can be interpreted as standardized coefficients.

Figure 1 illustrates both the one-factor and two-factor models, indicating the loadings on the latent factors. Model fit was estimated using various goodness-of-fit indices (see Table 3). A nonsignificant χ^2 value suggests a good model fit. Furthermore, the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA) were consulted. Values of CFI > .95 and RMSEA < .08 indicate close fit for small sample sizes (N < 250). Furthermore, a smaller Akaike Information Criterion (AIC) indicates which of the two models, which differed in complexity, fit the data better (see Bühner, 2008, for information about model fit). Table 3 compares the model fit of the two models under study.

Table 3. Fit Indices for the One-Factor and Two-Factor Models concerning Children's Phonological Processing Skills and Linguistic Abilities

	One-factor model	Two-factor model
χ^2 (df)	7.73 (2)	1.23 (1)
p (X^2)	.02	.27
CFI	.987	.999
RMSEA	.07	.02
AIC	13643.68	13639.18

As indicated by the χ^2, RMSEA, and CFI as well as by AIC, the two-factor model showed a better fit compared to a simple one-factor model. In addition, a χ^2 difference test favored the two-factor model ($\Delta\chi^2 = 6.5$, $\Delta df = 1$, $p < .05$). This result suggests that the two-factor model, which differentiates phonological processing skills from linguistic abilities, is comparatively more compatible with the data structure than a global model of language competence. Thus, our data support the assumption that phonological processing skills and linguistic abilities are separable in early preschool-age children. However, as hypothesized, the two latent factors were highly correlated ($r = .80$), thus reflecting the high correlative association between phonological memory and the linguistic measures of vocabulary and grammar (see Table 2).

How is reading literacy in Grade 2 predicted by early phonological processing skills and linguistic abilities? In a next step, we analyzed whether early indicators of phonological processing skills would indeed be more strongly associated with later basic reading skills, whereas linguistic abilities (grammatical and lexical knowledge) would have a stronger impact on reading comprehension. Although linguistic abilities and phonological processing skills were found to be better described as two separable dimensions than a global dimension of general language competence, when trying to specify a model to predict reading literacy through the factors of Linguistic Abilities and Phonological Processing within a single model, suppression effects were found. This is probably due to the fact that the two factors were highly correlated. Thus, we tested single models to compare the impact of linguistic abilities and phonological processing on later reading literacy. Specifically, we hypothesized that linguistic abilities and phonological processing skills would differ with respect to their impact on later reading literacy.

The correlations already presented in Table 2 show that basic reading skills, specifically reading speed and reading comprehension, are highly correlated in Grade 2. This is expected because – at least in the early school years – basic reading skills are a necessary precondition for reading comprehension. Thus, restrictions in basic reading skills may hinder children's reading comprehension. Therefore, we specified two models, one for phonological processing skills and a second model for linguistic abilities in which basic reading skills were accounted for when predicting reading comprehension. With respect to phonological processing skills and linguistic abilities, latent variables were modeled as in the CFA reported above. From these latent factors, a direct path to basic reading skills (SLS; reading speed) and to reading comprehension (ELFE; text comprehension) was indicated (see Figure 2). Again, missing data were adjusted with the full-information-maximum-likelihood (FIML) approach. Figure 2 illustrates the four models with standardized beta weights. Both models show good to very good model fit (for linguistic abilities: χ^2 = 2.91, *df* =1, *p* = .09; CFI = 1.0; RMSEA = .06; for phonological processing: χ^2 = 1.16, *df* =1, *p* = .28; CFI = 1.0; RMSEA = .02).

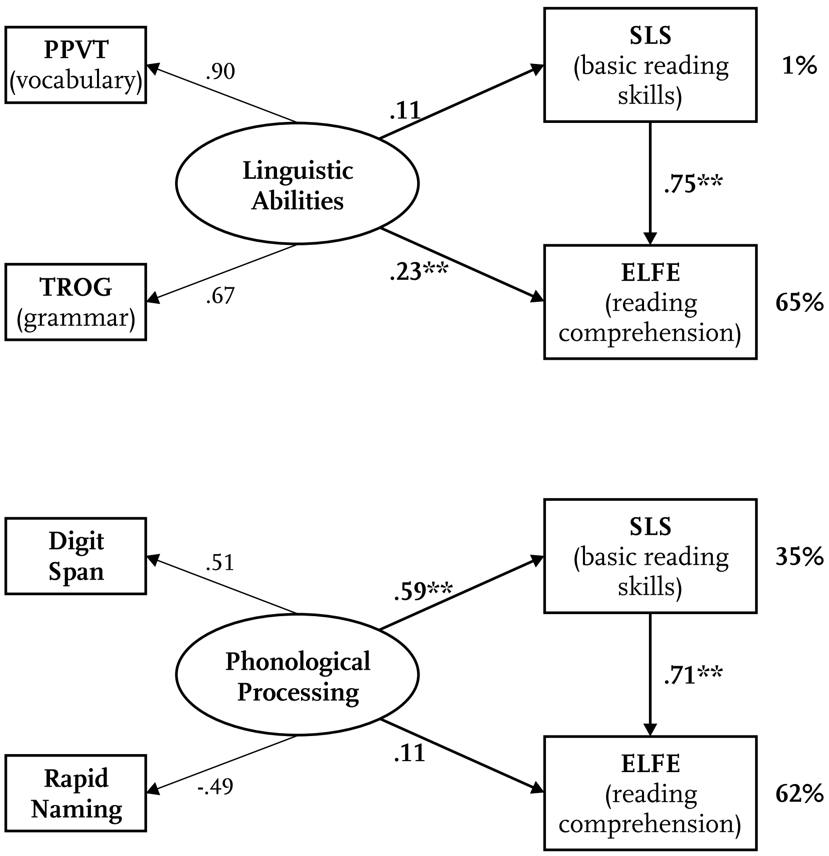

Figure 2. Predicting reading literacy (Grade 2) through phonological processing skills (Model 1) and linguistic abilities (Model 2). Circles represent latent variables and rectangles represent observed variables. All values can be interpreted as standardized coefficients.
$^{**}p < .01$.

As Figure 2 shows, linguistic abilities had a significant direct impact on reading comprehension, even when individual differences in basic reading skills were accounted for. In addition, an indirect effect of linguistic abilities on reading comprehension through basic reading skills was found (β = .18, p < .01). A direct link from linguistic abilities to basic reading skills did not appear.

By contrast and as hypothesized, phonological processing skills did not have a significant impact on reading comprehension when basic reading skills were controlled. Instead, phonological processing skills showed a strong impact on basic reading skills. This suggests that linguistic abilities (in the sense of the semantic and grammatical components of language) are specifically relevant for reading comprehension, even in Grade 2, whereas phonological processing skills indexed by phonological memory and the ability to quickly access lexical knowledge are particularly relevant for acquiring basic reading skills. However, phonological processing has an indirect effect on reading comprehension through basic reading skills.

Looking at the proportion of variance explained by the models, most of the variance in reading comprehension was explained by basic reading skills, which in turn were influenced by phonological processing skills.

Focus 2: Linguistic Competencies as Predictors of Later Reading Comprehension

Are there differential effects of early vocabulary and grammar on reading comprehension in Grade 2? As suggested by other studies as well as by our analyses, linguistic abilities that refer to the semantic and grammatical components of language have a specific significant impact on reading comprehension. This is true even with respect to early reading comprehension and when language predictors are assessed early in preschool. In a next step, we analyzed whether receptive vocabulary and grammar would each explain unique proportions of variance in reading comprehension or whether the variance shared between the two components would be relevant for reading comprehension. The goal of these analyses was to provide information about the relative impact of preschool children's early vocabulary and grammar as prerequisites for reading comprehension. Therefore, we conducted hierarchical regression analyses to explain the variance in reading comprehension. To determine the unique contributions of grammar and vocabulary, we conducted two hierarchical regression analyses. In Model A, vocabulary was entered in the first step and grammar in the second step; in Model B, the order was reversed. The increase of explained variance in the second step thus provides information about the unique contribution of the second predictor. The amount of shared variance can be determined by subtracting the independent contributions of the two predictors from the total explained variance. Furthermore, we controlled for basic reading skills as we did not predict reading comprehension per se but the residual variance of reading comprehension. Table 4 shows the results of the two hierarchical regression analyses.

Table 4. Summary of Hierarchical Regression Analyses Predicting the Residuum of Reading Comprehension (Controlling for Basic Reading Skills) from Vocabulary and Grammar

	β	t	R^2	ΔR^2
Model A				
Step 1				
Vocabulary (PPVT), time 3	.32	4.67	.10	
Step 2				
Grammar (TROG), time 3	.00	0.05	.10	.00
Model B				
Step 1				
Grammar (TROG), time 3	.17	2.33	.03	
Step 2				
Vocabulary (PPVT), time 3	.32	3.98	.10	.07**

Note. N = 193; PPVT = Peabody Picture Vocabulary Test; TROG = Test for the Reception of Grammar.
** $p < .01$

As Table 4 shows, when predicting the residuum of reading comprehension through grammar and vocabulary, only vocabulary explained specific variance. Vocabulary accounted for an additional 7% of the variance over and above the impact of grammar, $\Delta R^2 = .07$; $F_{inc}(1, 190) = 15.86$, $p < .01$. The unique contribution made by grammar to reading comprehension was zero, and even if entered in the first step, the amount of variance explained by grammar was small. Together, vocabulary and grammar at the ages of about 4 to 5 years explained 10% of the residual variance in reading comprehension in the second grade (i.e., more than 3 years later). Because the inclusion of grammar in a second step did not explain any additional variance, the amount of variance shared between grammar and vocabulary when predicting reading comprehension was 3%.

Do integrative language measures explain differences in later reading comprehension better than vocabulary and grammar? To answer our research question regarding whether more functional and integrative measures of language competencies would be able to predict later reading comprehension over and above the impact of vocabulary and grammar, data from the subsample of children who received the tests for story comprehension and reproduction as well as for sentence reproduction were considered. Table 5 shows the correlations between the various language measures and reading literacy (reading comprehension and basic reading skills) for this subsample. As in the whole sample, grammar and vocabulary were more strongly related to reading comprehension than to basic reading skills. Furthermore, story comprehension was significantly correlated with both kinds of linguistic abilities (i.e., grammar and vocabulary), whereas story reproduction was associated only with vocabulary, but not with grammar. For story comprehension and reproduction, their correlations with phonological processing skills (phonological memory and access to long-term memory) were small and even nonsignificant for rapid naming. Sentence reproduction, by contrast, was significantly related to all language measures. Furthermore, sentence reproduction showed higher correlations with phonological memory than any of the other language measures.

Interestingly, although story reproduction and comprehension were associated with the various language measures, significant correlations with either reading comprehension or basic reading skills were not found. By contrast, sentence reproduction was significantly related to both reading comprehension and basic reading skills. Again, as was found for the total sample, the correlation between reading comprehension and basic reading skills was particularly high.

Table 5. Correlations between Measures of Phonological Processing, Linguistic Abilities, and Reading Comprehension for the Subsample that was Tested on Story Reproduction and Comprehensions as well as on Sentence Reproduction

	1	2	3	4	5	6	7	8
1. Digit Span, time 3								
2. Rapid Naming, time 3	-.13							
3. PPVT, time 3	.36**	-.10						
4. TROG, time 3	.26**	-.13	.47**					
5. Story Reproduction, time 2	.23*	.00	.45**	.18				
6. Story Comprehension, time 2	.29**	-.15	.49**	.49**	.44**			
7. Sentence Reproduction, time 3	.45**	-.22*	.49**	.30**	.30**	.49**		
8. ELFE 1-6, Grade 2	.45**	-.19	.33**	.32*	.01	.17	.43**	
9. SLS 1-4, Grade 2	.36**	-.31**	.14	.22	.03	.12	.37**	.82**

Note. Correlations between rapid naming and the other measures are negative because the score on the measure is the time needed to complete the task. PPVT = Peabody Picture Vocabulary Test; TROG = Test for the Reception of Grammar; ELFE = Ein Leseverständnistest für Erst- bis Sechstklässler (reading comprehension); SLS = Salzburger Lesescreening für die Klassenstufen 1-4 (reading speed).
* $p < .05$. ** $p < .01$.

To determine whether integrative language competencies (i.e., more integrative and functional measures of language competencies that require the mastery and interplay of various language components) would explain independent proportions of variance over and above grammar and vocabulary, we again conducted hierarchical regression analyses. In a first step, children's grammatical and vocabulary knowledge were entered into the model. In a second step, measures that assessed integrative language skills were added to determine the specific variance explained by these measures over and above vocabulary and grammar. Again, we predicted the residuum of reading comprehension while controlling for basic reading skills.

Story comprehension and reproduction. Table 6 presents the results for the hierarchical regression analyses predicting the residuum of reading comprehension from individual differences in early vocabulary, grammar, and story comprehension and production.

Table 6. Summary of Hierarchical Regression Analyses Predicting the Residuum of Reading Comprehension (Controlling for Basic Reading Skills) from Vocabulary, Grammar, and Story Reproduction and Comprehension

	β	t	R^2	ΔR^2
Step 1				
Vocabulary (PPVT), time 3	.18	1.20		
Grammar (TROG), time 3	.21	1.43	.10	
Step 2				
Vocabulary (PPVT), time 3	.24	1.47		
Grammar (TROG), time 3	.18	0.98		
Story Reproduction, time 2	-.19	-1.14		
Story Comprehension, time 2	.12	0.67	.13	.03

Note. N = 52; PPVT = Peabody Picture Vocabulary Test; TROG = Test for the Reception of Grammar.

Table 6 shows that the integrative language measures (story reproduction and story comprehension) accounted for only a small amount of additional variance, $\Delta R^2 = .03$; $F_{inc}(1, 47) = 0.71$, *ns*, when individual differences in vocabulary and grammar were controlled. For story reproduction, a nonsignificant negative regression weight was obtained. This may be due to suppression effects as story reproduction was not correlated with reading literacy (cf. Table 5) but was correlated with vocabulary (see Bühner & Ziegler, p. 686). Thus, story reproduction and comprehension, although recognized as ecologically valid integrative language measures, did not explain additional variance over and above measures of language components (vocabulary, grammar) and were, in fact, only weakly associated with reading literacy in Grade 2.

Sentence reproduction. Table 7 presents the results of the hierarchical regression analyses predicting the residuum of reading comprehension from vocabulary, grammar, and sentence reproduction.

Table 7. Summary of Hierarchical Regression Analyses Predicting the Residuum of Reading Comprehension (Controlling for Basic Reading Skills) from Vocabulary, Grammar, and Sentence Reproduction

	β	t	R^2	ΔR^2
Step 1				
Vocabulary (PPVT), time 3	.17	1.09		
Grammar (TROG), time 3	.20	1.32	.10	
Step 2				
Vocabulary (PPVT), time 3	.16	1.05		
Grammar (TROG), time 3	.17	1.04		
Sentence Reproduction, time 3	.08	0.51	.10	.00

Note. N = 49; PPVT = Peabody Picture Vocabulary Test; TROG = Test for the Reception of Grammar.

As can be seen in Table 7, the addition of sentence reproduction in the second step did not improve the amount of variance explained by vocabulary and grammar. Thus, although sentence reproduction is a measure that draws on the child's lexical and grammatical knowledge as well as on his or her phonological information processing skills, it did not explain additional variance in the residuum of reading comprehension over and above measures of language components (vocabulary, grammar).

Thus, both integrative language indicators did not show additional effects on reading comprehension over and above the language components. To further substantiate this result, we conducted two additional analyses. Specifically, we predicted reading comprehension without controlling for basic reading skills (see Table 8).

Table 8. Summary of Hierarchical Regression Analyses Predicting Reading Comprehension in Grade 2 from Language Measures at Time 3 of the BiKS-3-10 Study

	β	t	R^2	ΔR^2
Model: Story Reproduction & Comprehension				
Step 1				
Vocabulary (PPVT), time 3	.24	1.84		
Grammar (TROG), time 3	.24	1.81	.16	
Step 2				
Vocabulary (PPVT), time 3	.37*	2.53		
Grammar (TROG), time 3	.26	1.84		
Story Reproduction, time 2	-.27	-1.80		
Story Comprehension, time 2	.01	0.03	.21	.05
Model: Sentence Reproduction				
Step 1				
Vocabulary (PPVT), time 3	.23	1.67		
Grammar (TROG), time 3	.24	1.78	.15	
Step 2				
Vocabulary (PPVT), time 3	.09	0.59		
Grammar (TROG), time 3	.19	1.42		
Sentence Reproduction, time 3	.32*	2.27	.23	.08*

Note. N = 61/57; PPVT = Peabody Picture Vocabulary Test; TROG = Test for the Reception of Grammar.
*p < .05

As Table 8 shows, story reproduction and comprehension did not account for a significant amount of additional variance, $\Delta R^2 = .05$; $F_{\text{inc}}(1, 56) = 1.85$, *ns*. Once again, a negative regression weight for story reproduction was obtained. This, once again, is potentially due to suppression effects. Story comprehension and reproduction seem to absorb variance from vocabulary and grammar that is not relevant for reading comprehension (see Bühner & Ziegler, p. 686). Thus, even when reading comprehension was considered instead of the residuum of reading comprehension, story comprehension and reproduction did not explain specific variance over and above the grammatical and lexical components of language.

By contrast, when sentence reproduction was entered in a second step after controlling for differences in vocabulary and grammar (see Table 8; Model Sentence Reproduction), sentence reproduction significantly improved the amount of variance explained in reading comprehension, $\Delta R^2 = .08$; $F_{\text{inc}}(1, 53) = 5.14$, $p < .05$.

Furthermore, the regression weights for vocabulary and grammar decreased when sentence reproduction was considered in the same analysis. Thus, sentence reproduction was found to be the comparatively strongest predictor of reading comprehension. This is the case most likely because sentence reproduction draws on both lexical and grammatical knowledge as well as on phonological processing skills. To further analyze the effect of sentence reproduction on reading comprehension, an additional analysis was conducted. We tested whether the effect of sentence reproduction would be mediated through phonological processing skills or whether it would have an effect over and above phonological processing. In this analysis, besides vocabulary and grammar, phonological memory and rapid automatized naming (RAN) were entered in a first step, and sentence reproduction was added in a second step. Table 9 shows the results of this analysis.

Table 9. Summary of Hierarchical Regression Analyses Predicting Reading Comprehension from Vocabulary, Grammar, Digit Span, and Sentence Reproduction

	β	t	R^2	ΔR^2
Step 1				
Vocabulary (PPVT), time 3	.14	1.40		
Grammar (TROG), time 3	.17	1.22		
Rapid Naming, time 3	-.15	-1.18		
Digit Span, time 3	.32	2.42	.23	
Step 2				
Vocabulary (PPVT), time 3	.09	0.63		
Grammar (TROG), time 3	.16	1.13		
Rapid Naming, time 3	-.14	-1.09		
Digit Span, time 3	.25	1.62		
Sentence Reproduction, time 3	.14	0.84	.24	.01

Note. $N = 54$; The obtained β weights for rapid naming are negative because the score on the measure is the time needed to complete the task. PPVT = Peabody Picture Vocabulary Test; TROG = Test for the Reception of Grammar.

As Table 9 shows, the unique variance that sentence reproduction explained was indeed due to the variance shared between sentence reproduction and phonological processing skills. Sentence reproduction did not contribute further to the prediction of reading comprehension when differences in the grammatical and lexical components of language as well as phonological processing skills were statistically controlled, $\Delta R^2 = .01$; $F_{inc}(1, 48) = 0.70$, *ns*.

The results of these analyses also demonstrate that early language measures account for a higher proportion of variance in reading comprehension when basic reading skills are not accounted for. Although reading comprehension was assessed 3 to 4 years later than oral language competencies, language measures explained up to 24% of the variance in reading comprehension in Grade 2.

Discussion

The aim of the present study was to investigate how various language indicators assessed early in preschool would predict reading literacy in the first years of formal reading instruction in primary school. We tested and confirmed that phonological processing skills (phonological working memory, fast access to phonological representations in long-term memory) and linguistic abilities (vocabulary, grammar) are significantly interrelated in preschool-age children but nevertheless contribute in different ways to the development of early reading literacy – that is, to the acquisition of basic reading skills (reading speed) and reading comprehension, respectively. In addition, we analyzed the specific long-term impact of early individual differences in vocabulary, grammar, and integrative language measures (story reproduction and comprehension, sentence reproduction) in preschool on reading comprehension. Our study indicated that when language components (grammar, vocabulary) were considered together with integrative language measures, the latter did not explain an additional or higher amount of variance in early reading comprehension. In the following, the main results of the study will be discussed in more detail and related to other research outcomes.

Based on theoretical models and empirical results concerning precursors and predictors of reading literacy, we first analyzed whether the distinction between phonological processing skills and linguistic abilities as predictors of different facets of reading literacy could be empirically substantiated in the early preschool years. Confirmatory factor analyses demonstrated that a two-factor model that differentiated between phonological processing skills and linguistic abilities fit the data better than a one-factor model that integrated all language measures into one global factor. Thus, the distinction between phonological processing skills and linguistic abilities seems empirically justified. However, it should be kept in mind that the estimated correlation between the latent factors of phonological processing and linguistic abilities was strong (*r* = .80), reflecting and substantiating the proposed tight developmental interrelations between various language facets (e.g., vocabulary acquisition and phonological working memory; Ebert, et al., 2013; Gathercole, et al., 1992; Weinert, et al., 2012). In fact, interindividual differences in digit span as an indicator of phonological working memory capacity were even more highly correlated with grammar and vocabulary than with rapid naming as an indicator of the fast access to phonological representations in long-term memory. Thus, correlational analyses showed that digit span and rapid naming as indicators of phonological processing skills are not more strongly interconnected with each other than each of these indicators is related to vocabulary and grammar as indicators of linguistic abilities. However, vocabulary and grammar were more strongly connected to each other than to phonological working memory (i.e., digit span, in this case). This result suggests that the linguistic measures (vocabulary and grammar) may refer to the same underlying construct or have similar developmental determinants, whereas digit span and rapid naming, although related, may be connected to this construct for other developmental reasons.

One might object that we didn't assess measures of phonological awareness as an important facet of phonological information processing, which seems to be one of the most important predictors of reading development (e.g., Shanahan & Lonigan, 2010). Admittedly, as already mentioned, phonological awareness seems to be less important in languages with comparatively more regular orthography (e.g., German). Moreover, the empirical data suggest that measures of phoneme skills (also known as phonological awareness in a narrow sense) show floor effects when assessed at such an early age as in the present study, whereas measures of onset-rime skills (also known as phonological awareness in a broader sense) are often not associated with early reading development and have been shown to be more strongly correlated with vocabulary than with phoneme awareness (Muter, et al., 2004). These results are also in line with findings from the BiKS-3-10 study that are not reported in the results section: For a subgroup of children, a measure of rhyming was assessed at a later time point than the measures reported here. Confirmatory factor analyses (similar to those conducted in the present study) including the rhyming task found rhyming to be more strongly associated with vocabulary than with the other measures of phonological awareness.

Thus, it cannot be argued that a confirmatory factor analysis including additional measures of phonological awareness to represent phonological processing skills would have produced a clearer distinction between phonological processing skills and linguistic abilities. Furthermore, a study conducted by Lonigan et al. (2009) that did not consider linguistic skills but only aspects of phonological processing showed that phonological awareness was more closely related to phonological working memory than to rapid automatized naming. Specifically, a two-factor model combining measures of phonological awareness and phonological memory in one factor and measures of rapid automatized naming in a second factor fitted the data best. Thus, this study also demonstrates an exceptional position of rapid automatized naming, whereas phonological awareness and phonological working memory seem to be more strongly associated. These results may also excuse the fact that we did not consider measures of phonological awareness. Actually, with respect to our second aim, a strength of the present study is that we considered phonological working memory and rapid automatized naming as indicators of phonological processing skills.

The second aim of the present study was to replicate the finding that lexical and grammatical knowledge are especially relevant to reading comprehension, whereas phonological processing skills are more important for basic reading skills. To date, only a few studies have considered various aspects of phonological processing as well as of linguistic abilities within one and the same study (Cain, 2010). Also, if both facets were included, they most often focused on phonological awareness, but not on other measures of phonological information processing (e.g., Muter, et al., 2004; Senechal, et al., 2006). The study by Muter et al. (2004), for example, assessed children's vocabulary and grammar as we did in the present study, whereas phonological processing was indicated by phonological awareness (onset-rime and phoneme awareness); measures of phonological memory and rapid automatized naming were not considered. Thus, the present study was able to provide new information by verifying that aspects of phonological information processing other than phonological awareness show similar effects on later reading and different effects than linguistic abilities. Muter et al. (2004) showed that measures of phoneme awareness (but not of onset-rime) at age 5 were significant predictors of word recognition at age 6 even when word recognition at age 5 and early vocabulary and grammar were controlled for, whereas vocabulary and grammar failed to predict the growth of word recognition over and above phoneme awareness. With respect to reading comprehension, the pattern was reversed: Vocabulary and grammar showed an effect on reading comprehension but not on phoneme awareness when earlier word recognition was controlled. The results of the present study were pretty much the same, although we used digit span and rapid automatized naming as indicators of phonological processing instead of measures of phoneme awareness and a measure of sentence comprehension for grammar instead of a word-order correction task and a morphological generation task. The present study showed that phonological processing skills assessed at the age of 4 had a significant effect on basic reading skills (reading speed/fluency) in Grade 2 (at the age of about 8 years), whereas no effect of phonological skills on reading comprehension showed up after controlling for basic reading skills. By contrast, linguistic abilities asserted a significant effect on reading comprehension after controlling for basic reading skills, but not on basic reading skills. Thus, the present study demonstrates that, no matter what aspects of phonological processing are assessed and even when phonological awareness is not considered, phonological processing is more important for basic reading skills, whereas linguistic abilities are specifically relevant for reading comprehension. Moreover, the present study further shows that this pattern of results holds (a) when predictors are assessed at a very young age (4 years) and (b) for early reading literacy (i.e., in a developmental phase when reading comprehension may still be dominated and restricted by decoding processes). Indeed, correlations between basic reading skills and reading comprehension are high in children in Grade 2. In this context, a meta-analysis conducted by Gough, Hoover, and Peterson (1996, cited by Muter, et al., 2004) demonstrated that the correlations between decoding and reading comprehension are high in the early grades but decrease later on. Thus, although basic reading skills and reading comprehension are highly redundant indicators of reading literacy in the early years of reading instruction, our results demonstrate that there are already important differences concerning the relevance of various language predictors. This result shows that basic reading skills and reading comprehension have different determinants in development from early on (see also Cain & Oakhill, 2007).

Furthermore, it should be noted that we found significant effects of language predictors on reading literacy over a long time period of 4 years (note that this is half of these children's lifetimes) including the transition from one learning environment (preschool) to a new learning environment (school). Thus, during this time, large environmental influences on reading development are to be expected, and these could have obscured or reduced the impact of variables measured in preschool. Nevertheless, in this study as well as in others, child variables were found to be strong predictors of developmental progress (Ebert, et al., 2013). To be sure, these developing child variables are – in accordance with bioecological models of development (Bronfenbrenner & Morris, 2006) – influenced by each other as well as by environmental factors (Weinert & Ebert, 2013).

The third aim of the present study was to investigate in more detail the predictive power of linguistic abilities (vocabulary, grammar) and more integrative language measures on reading comprehension. In comparison to phonological processing and its relevance to more basic reading skills (decoding, reading fluency), relatively less is known about the relative impact of various linguistic abilities on reading comprehension (Cain, 2010; Muter, et al., 2004). First, we analyzed whether vocabulary or grammar would have a comparatively stronger impact on later reading comprehension when considered simultaneously. Whereas most studies have considered just vocabulary and not grammar, a study by Muter et al. (2004) demonstrated that the impact of grammar and vocabulary on reading comprehension was quite similar. By contrast, the present study showed that vocabulary but not grammar (sentence comprehension) at the age of 4 explained additional variance in reading comprehension in Grade 2 after controlling for basic reading skills (reading speed) and grammar or vocabulary respectively. In addition, a study by Roth et al. (2002) showed that semantic abilities assessed in kindergarten more strongly predicted reading comprehension than a test of syntax.

From a theoretical point of view, predictions concerning the relative importance of vocabulary and grammar are not straightforward. Obviously, their relative impacts may depend on features of the written text (complexity of sentence structure and vocabulary), the assessment of text comprehension (e.g., the extent to which it taps one or the other aspect; the extent to which it presupposes specific processes of text comprehension), as well as on the time point of assessment in language and reading development (When are the linguistic predictors assessed in preschool? When is reading literacy assessed in school?). In order to comprehend (written) texts, the child has to have both lexical and grammatical knowledge; in addition these two language components are interrelated in language development and may interact in text comprehension (see e.g., Weinert, 2006). Emanating from the fact that linguistic skills are relatively stable across time (e.g., Storch & Whitehurst, 2002; Weinert, Ebert, & Dubowy, 2010), reading comprehension may depend on whether texts include complex grammar and/or complex vocabulary. For example, when texts include more complex vocabulary, more sophisticated grammatical abilities may help the reader to construct the meaning of the text even without knowing all the words, whereas if sentence and text structure draw on basic linguistic skills, it may be sufficient to know most of the words to make sense of the text. Thus, the relative impact of vocabulary and grammar in predicting reading may change according to text complexity and/or a child's age. Oakhill, Cain, and Bryant (2003, cited in Cain & Oakhill, 2007), for example, found that syntactic ability did not predict reading comprehension in 7- or 8-year-olds when controlling for differences in vocabulary and IQ, but predicted reading comprehension 1 year later. Thus, the more dominant impact of vocabulary found in our study may be due to the fact that (written) texts are still rather easy in second grade when taking into account the still restricted basic reading competencies of the children. Interestingly, further analyses of our data showed that, at later time points and regarding subgroups of older children, grammar and vocabulary explained a similar amount of specific variance. This result is in line with the above-mentioned study by Muter et al. (2004). This suggests that the relative predictive power of vocabulary and grammar might also change according to children's age. Because one possible explanation for the diverging results traces back to the developmental relation between vocabulary and grammar, future research should consider the developmental pathways between these language variables in more detail.

Besides the issue of the relative importance of specific language components (e.g., vocabulary and grammar) for reading literacy, another aim of the present study was to further investigate whether measures of more integrative and functional language competencies would be better predictors of reading comprehension than measures of language components. Integrative language measures such as oral text comprehension or sentence reproduction require not only the availability of lexical and grammatical knowledge, but also tap phonological processing skills and are more ecologically valid.

Concerning oral text comprehension (listening comprehension), van den Broek et al. (2005) found strong connections between 4-year-olds' listening comprehension and their reading comprehension in Grade 2. This was true for important causal information that the children remembered in free recall (r = .58) as well as for their scores on complex questions (r = .53). Even after controlling for vocabulary as well as for letter and word identification and phonemic awareness, the predictive power of oral text comprehension in preschool for reading comprehension in the second grade remained significant. These results are in contrast to those of the present study. Story reproduction and story comprehension at the age of about 4 years did not explain additional variance in reading comprehension after controlling for differences in basic reading skills (reading speed), vocabulary, and grammar. Thus, these integrative measures did not show a predictive effect over and above measures of language components. Even the simple correlations between our measures of oral text comprehension and reading literacy were small. With respect to story comprehension, correlations ranged between r = .12 and r = .17; when considering story reproduction, they were almost zero. How can we explain these diverging results? One explanation may be found in our operationalization of oral text comprehension. Van den Broek et al. (2005), for example, differentiate between various types of "comprehension" (e.g., the ability to remember information explicitly given in the text, to apply information conveyed in the text, to recognize the topic or moral of a text, or to provide a critical appraisal of the text). However, according to van den Broek and colleagues, these types of comprehension share core processes that "involve interpretation of the information in the text, the use of prior knowledge to do so and, ultimately, the construction of a coherent representation or picture what the text is about" (van den Broek, et al., 2005, p. 109). Similar to this approach, our comprehension measures included indicators of the information remembered by the child and of the inferences drawn. However, it is possible that our measure is more dependent on memory resources than other measures of oral text comprehension because the presentation of the story and the assessment of story comprehension (story reproduction, comprehension questions) were separated by a retention interval. Consistent with this explanation, Reese et al. (2010) found that the quality of narratives (in contrast to, e.g., story memory) was most predictive of reading. Unfortunately, we do not know what kind of comprehension questions van den Broek and colleagues used in their study; however, because the children in van den Broek's study listened to an extended orally presented or televised narrative, it is reasonable to assume that they had to answer mainly comprehensive questions about the causal structure of the story. By contrast, in our task, children had to remember a greater number of details from a relatively short story.

However, our results are puzzling because our measures of oral text comprehension were significantly correlated with vocabulary, grammar, and sentence reproduction (except for the low correlation between grammar and story reproduction), but not with reading literacy. Oral text comprehension and other language measures are related to each other and may interact with each other; thus, linguistic skills in the sense of vocabulary or grammar may limit children's comprehension skills. Nevertheless, contrary to our expectations and in contrast to other research results (e.g., de Jong & van der Leij, 2002; van den Broek et, al., 2005), we did not find that oral text comprehension in preschool was an important predictor of later reading comprehension in Grade 2. However, we do not know whether this is due to the measures of oral text comprehension or to the measure of reading comprehension administered in the present study. Thus, our test of reading comprehension may tap more basic rather than linguistically challenging comprehension skills. This is usually the case in the early stages of children's reading development as children are still struggling with basic reading skills. Accordingly, as already mentioned, reading comprehension and basic reading skills were highly correlated in our study (i.e., even after 2 years of reading instruction).

Another important contribution of the present study is that, over and above considering the impact of language components (vocabulary, grammar) as well as of more integrative, functional language measures (story comprehension, story reproduction) on reading comprehension, we also introduced a second measure of integrative language competencies, namely, sentence reproduction. This measure is of specific interest because it is supposed to be a highly reliable predictor of reading development (Goldammer, et al., 2010). Compared to oral text comprehension, it is conceptually less similar to reading comprehension and not as ecologically valid but easy to assess. In particular, sentence reproduction comprises various language skills known to be predictive of basic reading skills and reading comprehension. On the one hand, sentence reproduction draws on phonological processing skills to verbally store the presented sentence in short-term memory; on the other hand, available lexical and grammatical knowledge support these memory processes as well as the immediate reconstruction of the semantic and grammatical structure of the given sentence. Therefore, we tested whether this measure significantly predicts reading comprehension and outperforms indicators of language components. As an extension of the study by Goldammer et al. (2010), we examined the impact of a sentence reproduction task on reading comprehension instead of on basic reading skills. Contrary to the results of Goldammer et al. (2010), who found sentence reproduction at the age of 5 years to be the strongest predictor of basic reading skills (reading speed at the word and sentence levels) at about 8 years, our results showed that a higher attainment in sentence reproduction at age 4 did not explain unique variance in children’s reading comprehension in Grade 2 over and above language components, that is, after controlling for differences in basic reading skills (reading speed at the sentence level), vocabulary, and grammar. However, if we did not control for basic reading skills, sentence reproduction was a stronger predictor than vocabulary and grammar. This seems to be due to the fact that sentence reproduction draws on language components as well as on phonological processing skills, specifically verbal memory. This assumption is supported by the finding that sentence reproduction did not explain additional unique variance in reading comprehension when individual differences in grammar, vocabulary, verbal short-term memory (digit span), and fast access to long-term memory (rapid automatized naming) were statistically controlled. Thus, our results suggest that sentence reproduction is a highly valid predictor of reading comprehension because of its demands on phonological processing and linguistic abilities. Accordingly, we recommend that researchers use sentence reproduction as an economical measure of children’s general language competencies. However, this measure does not assess (language) competencies over and above the required language components (i.e., tests of language components). Furthermore, our results suggest that the predictive effect of sentence reproduction on reading comprehension is mediated mainly through its interrelation with basic reading skills.

Thus, sentence reproduction does not account for unique variance (over and above vocabulary and grammar) in reading comprehension when basic reading skills (reading speed) are controlled. This converges with the results of Goldammer and colleagues, who found that sentence reproduction was a strong predictor of basic reading skills.

In sum, our results are in line with the national and international research literature on the impact of phonological processing skills and linguistic abilities on the development of reading literacy. Specifically, we replicated the differential effects in the predictive power of phonological processing skills and linguistic abilities. Whereas phonological processing skills are superior predictors of basic reading skills, linguistic abilities are more important in the prediction of reading comprehension when controlling for basic reading skills that may hinder more complex comprehension processes from taking place (see also Muter, et al., 2004; Senechal, et al., 2006). Thus, phonological processing skills are important for reading comprehension as long as basic reading skills are not accounted for; when basic reading skills are controlled, phonological processing skills no longer account for reading comprehension. In the same vein, linguistic abilities are subordinate when basic reading skills are not controlled. In particular, the present study provides important new information as the differential effects of phonological processing skills and linguistic abilities on reading literacy even hold when predictors are assessed at an early age in preschool and for reading outcomes in early phases of reading development.

Probably the most important contribution of the present study is that it partly disentangles the differential effects of various oral language indicators on early reading comprehension. Specifically, analyses were conducted with respect to the impact of language components (vocabulary, grammar) compared to integrative language competencies (oral text comprehension, sentence reproduction). This is especially important when thinking about the promotion of oral language in preschool. Our results suggest that early in preschool, it is vocabulary in particular that seems to provide the foundation for further language and reading development. However, it is likely that vocabulary influences grammar and integrative language skills such as oral text comprehension, which may become more important in the course of development when children grow older. Concerning integrative language competencies, our results are somewhat contradictory to studies that found that oral text comprehension, even early in preschool, was a reliable predictor of reading comprehension. It is possible that this difference is due to our measures of oral text comprehension or to our measure of reading comprehension. In fact, there is a deep need for more reliable and valid tests of oral text (and reading) comprehension in young children. Further studies have to gain insight into the developmental interrelations of vocabulary, grammar, and more integrative language measures, which are all subject to social disparities from an early age (e.g., Ebert, et al., 2013; Weinert & Ebert, 2013; Weinert, et al., 2010, 2012). Because individual differences have been shown to be rather stable in the language domain, this might be important not only for reading comprehension but also for school learning in general.

References

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243-277). Mahwah, NJ: Erlbaum.

Berendes, K. (2011). *Relevanz und Therapie der phonologischen Informationsverarbeitung bei Lese-Rechtschreib-Schwierigkeiten in der orthographischen Phase des Schriftspracherwerbs*. Dissertation, Medizinische Hochschule Hannover.

Bishop, D. V. (1983/1989). *Test of Reception of Grammar (TROG)*. Manchester, U.K.: Medical Research Council.

Bradley, L., & Bryant, P. E. (1985). *Rhyme and reason in reading and spelling*. Ann Arbor, MI: University of Michigan Press.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Gen. Eds.), *Handbook of child psychology* (6th ed.). In R. M. Lerner (Vol. Ed.), Vol. 1: Theoretical models of human development (pp. 793-828). New York: Wiley.

Bryant, P. E., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26*, 429-438. doi:10.1037/0012-1649.26.3.429

Bühner, M., & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson Studium.

Byrne, B., & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young children. A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology, 87,* 488-503. doi:10.1037/0022-0663.87.3.488

Cain, K. (2010). *Reading development and difficulties: An introduction*. Oxford: Wiley-Blackwell.

Cain, K., & Oakhill, J. (2007). Reading comprehension difficulties: Correlates, causes, and consequences. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 41-75). New York, NY: Guilford Press.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91,* 77-111. doi:10.1016/S0010-0277(03)00164-1

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33,* 934-945. doi:10.1037/0012-1649.33.6.934

De Jong, P. F., & Leseman, P. P. M. (2001). Lasting effects of home literacy on reading achievement in school. *Journal of School Psychology, 39,* 389-414. doi:10.1016/S0022-4405(01)00080-2

De Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91,* 450-476. doi:10.1037/0022-0663.91.3.450

De Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading, 6,* 51-77. doi:10.1207/S1532799XSSR0601_03

Dickinson, D. K., Golinkoff, R. M., & Hirsh-Pasek, K. (2010). Speaking out for language: Why language is central to reading development. *Educational Researcher, 39,* 305-310. doi:10.3102/0013189X10370204

Dickinson, D. K., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E. S., & Poe, M. D. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology, 95,* 465-481. doi:10.1037/0022-0663.95.3.465

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test - Revised* (PPVT-R). Circle Pines, MN: American Guidance Service.

Ebert, S., Lockl, K., Weinert, S., Anders, Y., Kluczniok, K., & Rossbach, H.-G. (2013). Internal and external influences on vocabulary development in preschool age. *School Effectiveness and School Improvement, 24,* 138-154. doi:10.1080/09243453.2012.749791.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250-287. doi:10.1598/RRQ.36.3.2

Ennemoser, M., Marx, P., Weber, J., & Schneider, W. (2012). Spezifische Vorläuferfertigkeiten der Lesegeschwindigkeit, des Leseverständnisses und des Rechtschreibens. Evidenz aus zwei Längsschnittstudien vom Kindergarten bis zur 4. Klasse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 44,* 53-67. doi:10.1026/0049-8637/a000057

Fox, A. (2006). *Test zur Überprüfung des Grammatikverständnisses (TROG-D)*. Idstein: Schulz-Kirchner.

Ganzeboom, H. B. G., de Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21,* 1-56. doi:10.1016/0049-089X(92)90017-B

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28,* 200-213. doi:10.1016/0749-596X(89)90044-2

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28,* 887-898. doi:10.1037/0012-1649.28.5.887

Georgiou, G. K., Parrila, R. K., & Papadopoulus, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology, 100,* 566-580. doi:10.1037/0022-0663.100.3.566

Goldammer, A. von, Mähler, C., Bockmann, A.-K., & Hasselhorn, M. (2010). Vorhersage früher Schriftsprachleistungen aus vorschulischen Kompetenzen der Sprache und der phonologischen Informationsverarbeitung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42,* 48-56. doi:10.1026/0049-8637/a000005

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read.* Hillsdale, NJ: Erlbaum.

Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1-13). Mahwah, NJ: Erlbaum.

Grimm, H. (2001). *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5).* Göttingen. Hogrefe.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2,* 127-160. doi:10.1007/BF00401799

Kim, Y.-S., Wagner, R. K., & Foster, E. (2011). Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first-grade readers. *Scientific Studies of Reading, 15,* 338-362. doi:10.1080/10888438.2010.493964

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 95,* 453-464. doi:10.1037/0022-0663.95.3.453

Lenhard, W., & Schneider, W. (2006). *Ein Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6).* Göttingen: Hogrefe.

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology, 45,* 764-781. doi:10.1037/a0014132

Lonigan, C. J., Anthony, J. L., Phillips, B. M., Purpura, D. J., Wilson, S. B., & McQueen, J. D. (2009). The nature of preschool phonological processing abilities and their relations to vocabulary, general cognitive abilities, and print knowledge. *Journal of Educational Psychology, 101,* 345-358. doi:10.1037/a0013837

Lundberg, I., Frost, J., & Peterson, O.-P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263-284. doi:10.1598/RRQ.23.3.1

Lundberg, I., Larsman, P., & Strid, A. (2012). Development of phonological awareness during the preschool year: The influence of gender and socio-economic status. *Reading and Writing: An Interdisciplinary Journal, 25,* 305-320. doi:10.1007/s11145-010-9269-4

Mayringer, H., & Wimmer, H. (2003). *Salzburger Lese-Screening für die Klassenstufen 1-4 (SLS 1-4)*. Bern: Huber.

Melchers, P., & Preuss, U. (2003). *Kaufmann-Assessment-Battery for Children (K-ABC)*, German Version. Göttingen: Hogrefe.

Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., et al. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology, 90,* 347-373. doi:10.1037/0022-0663.90.3.347

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40,* 665-681. doi:10.1037/0012-1649.40.5.665

Muthén, L. K., & Muthén, B. O. (2010). *Mplus. Statistical analysis with latent variables.* Los Angeles, CA: Muthén & Muthén.

NICHD Early Child Care Research Network. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology, 41,* 428-442. doi:10.1037/0012-1649.41.2.428

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27,* 342-356. doi:10.1111/j.1467-9817.2004.00238.x

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. Language and Cognitive Processes, 18, 443-468. doi:10.1080/01690960344000008

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98,* 554-566. doi:10.1037/0022-0663.98.3.554

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Reese, E., Suggate, S., Long, J., & Schaughency, E. (2010). Children's oral narrative and reading skills in the first 3 years of reading instruction. *Reading and Writing: An Interdisciplinary Journal, 23,* 627-644. doi:10.1007/s11145-009-9175-9

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research, 95,* 259-272. doi:10.1080/00220670209596600

Roßbach, H.-G., Tietze, W., & Weinert, S. (2005). *Peabody Picture Vocabulary Test - Revised*. Unpublished German Research Version of the PPVT-R (L. M. Dunn & L. M. Dunn, 1981). Universität Bamberg, FU-Berlin.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 265-282. doi:10.1037/0022-0663.96.2.265

Schneider, W., Ennemoser, M., Roth, E., & Küspert, P. (1999). Kindergarten prevention of dyslexia: Does training in phonological awareness work for everybody? *Journal of Learning Disabilities, 32,* 429-437. doi:10.1177/002221949903200508

Sénéchal, M., & LeFevre, J.-A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73,* 445-460. doi:10.1111/1467-8624.00417

Sénéchal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 173-182). New York, NY: Guilford Press.

Shanahan, T., & Lonigan, C. J. (2010). The National Early Literacy Panel: A summary of the process and the report. *Educational Researcher, 39,* 279-285. doi:10.3102/0013189X10369172

Souvignier, E., Duzy, D., Glück, D., Pröscholdt, M. V., & Schneider, W. (2012). Vorschulische Förderung der phonologischen Bewusstheit bei Kindern mit Deutsch als Zweitsprache – Effekte einer muttersprachlichen und einer deutschsprachigen Förderung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 44,* 40-51. doi:10.1026/0049-8637/a000059

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86,* 24-53. doi:10.1037/0022-0663.86.1.24

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38,* 934-947. doi:10.1037/0012-1649.38.6.934

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities, 27,* 276-286. doi:10.1177/002221949402700503

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 579-593. doi:10.1037/0022-0663.91.4.579

Torppa, M., Lyytinen, P., Erskine, J., Eklund, K., & Lyytinen, H. (2010). Language development, literacy skills, and predictive connections to reading in Finnish children with and without familial risk for dyslexia. *Journal of Learning Disabilities, 43,* 308-321. doi:10.1177/0022219410369096 van den Broek, P., Kendeou, P., Kremer, K., Lynch, J., Butler, J., White, M. J., et al. (2005). Assessment of comprehension abilities in young children. In S. G. Paris & S. A. Stahl (Eds.), *Children´s reading comprehension and assessment* (pp. 107-130). Mahwah, NJ: Erlbaum.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192-212.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85,* 83-103. doi:10.1037/0022-0663.85.1.83

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., et al. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33,* 468-479. doi:10.1037/0012-1649.33.3.468

Weinert, S. (2006). Sprachentwicklung. In W. Schneider & B. Sodian (Eds.), *Kognitive Entwicklung* (Enzyklopädie der Psychologie C/V/2, pp. 609-719). Göttingen: Hogrefe.

Weinert, S. (2010a). Beziehungen zwischen Sprachentwicklung und Gedächtnisentwicklung. In H.-P. Trolldenier, W. Lenhard, & P. Marx (Eds.), *Brennpunkte der Gedächtnisforschung: Entwicklungs- und pädagogisch-psychologische Perspektiven* (pp. 147-170). Göttingen: Hogrefe.

Weinert, S. (2010b). Erfassung sprachlicher Fähigkeiten. In E. Walther, F. Preckel, & S. Mecklenbräuker (Eds.), *Befragung von Kindern und Jugendlichen* (pp. 227-262). Göttingen: Hogrefe.

Weinert, S., & Ebert, S. (2013). Spracherwerb im Vorschulalter: Soziale Disparitäten und Einflussvariablen auf den Grammatikerwerb. *Zeitschrift für Erziehungswissenschaft.* doi:10.1007/s11618-013-0354-8.

Weinert, S., Ebert, S., & Dubowy, M. (2010). Kompetenzen und soziale Disparitäten im Vorschulalter. *Zeitschrift für Grundschulforschung, 1,* 32-45.

Weinert, S., Ebert, S., Lockl, K., & Kuger, S. (2012). Disparitäten im Wortschatzerwerb: Zum Einfluss des Arbeitsgedächtnisses und der Anregungsqualität in Kindergarten und Familie auf den Erwerb lexikalischen Wissens. *Unterrichtswissenschaft, 40,* 4-25.

Weinert, S., & Grimm, H. (2008). Sprachentwicklung. In R. Oerter & L. Montada (Eds.), *Entwicklungspsychologie* (pp. 502-534). Weinheim: Beltz.

Wimmer, H. (1982). *Zur Entwicklung des Verstehens von Erzählungen*. Bern: Huber.

Wimmer, H. (2006). Don't neglect reading fluency! *Developmental Science, 9,* 447-448. doi:10.1111/j.1467-7687.2006.00527.x

Wimmer, H., Mayringer, H., & Landerl, K. (1998). Poor reading: A deficit in skill-automatization or a phonological deficit? *Scientific Studies of Reading, 2,* 321-340. doi:10.1207/s1532799xssr0204_2

Chapter 6

6 Social and Immigration-Specific Differences in the Development of Reading Comprehension: A Longitudinal Analysis of Primary School Students in Germany

Thorsten Schneider and Maximilian Pfost

Summary

According to the theory of social reproduction, parents' cultural habits, activities, and goods have large impacts on children's skills, knowledge, competencies, and educational attainment (Bourdieu, 1974; Bourdieu & Passeron, 1977). The cultural mobility model is less restrictive and less unidirectional than the theory of social reproduction. According to the cultural mobility model, students from lower social classes, in particular, can promote their school performance if they invest in cultural activities, thus attenuating the relation between their parents' class position and their own school success (Aschaffenburg & Maas, 1997; DiMaggio, 1982). In recent times, the school performance of students from immigrant families has been the focus of attention. Cultural capital is often context specific

Author Note

Thorsten Schneider,
Institute of Sociology, Universität Leipzig, Germany.

Maximilian Pfost,
Department of Educational Research, University of Bamberg, Germany.

This research was supported by grants BL 381/7-1, BL 381/7-2, and BL 381/7-3 from the German Research Foundation (DFG).

We thank Kristine Blatter (University of Bamberg) and Volker Stocké (University of Kassel) for their helpful comments on a draft of this chapter.

Correspondence concerning this chapter should be addressed to Thorsten Schneider, Institute of Sociology, Universität Leipzig, Beethovenstraße 15, 04107 Leipzig, Germany. E-mail: thorsten.schneider@uni-leipzig.de

and might lose its value as a consequence of immigration. Therefore, the relation between parents' cultural capital and students' school success should be weaker. However, according to the cultural mobility model, the relation between students' own cultural capital and school success should be stronger.

To provide new evidence on this topic, we analyzed panel data with value-added models on reading literacy from Grades 3 to 4. The data were derived from the BiKS-8-14 longitudinal study (Educational Processes, Competence Development, and Selection Decisions in Preschool- and School-Age Children) that have been collected in two German states since 2006.

Our empirical analysis on progress in reading showed that the gap in reading comprehension between students from families with low and high education increases across time. There is evidence that participation in highbrow culture fosters progress in reading comprehension, especially when parents participate in such activities. In addition, the amount of reading in which a student engages has a strong influence. However, no effects could be found for the amount of time parents read newspapers or books, the number of books at home, or children's use of libraries. Our results provide support for theories on social reproduction (strong influence of parents' education and highbrow activities), but are also consistent with an extended version of the cultural mobility model (the influence of students' reading habits). Most indicators of various forms of cultural capital have similar effects in native and immigrant families.

In the sociology of education, the concept of "cultural capital" has been intensively debated and used in research for explaining social inequality in educational attainment. The term was brought into sociology and familiar disciplines by Bourdieu (1974; Bourdieu & Passeron, 1977). Subsequent studies have been more or less connected to this concept. DiMaggio's (1982; DiMaggio & Mohr, 1985) contributions and re-interpretation in particular have been very influential in the English-speaking research community (for a review, see Lareau & Weininger, 2003). However, despite similarities, there are also substantial differences between these two concepts. Major discrepancies concern the breadth of cultural capital and the potential to make strategic investments in cultural capital to promote the upward mobility of offspring from less privileged families.

Stemming from the general concept of cultural capital, research has tried more and more to figure out which kind of cultural capital generates advantages in which educational areas. Researchers ask which aspects are relevant for the acquisition of cognitive competencies, which factors influence teachers' grading practices, and which kind of cultural capital is of special importance for the parent-teacher interaction. The main mechanisms that are discussed are socialization, cognitive stimulation, and signaling. In addition, a broad research strand has focused on reading habits. In sociology, this is often done under the headline of cultural capital (De Graaf, De Graaf & Kraaykamp, 2000; Sullivan, 2001). Current educational research is looking closer at the development and educational careers of children raised by immigrants, but little research has been conducted on the importance of cultural capital for educational success in immigrant families. As cultural capital is often context specific, it might be obliterated after a family immigrates to a new country.

In this chapter, we investigate the importance of cultural capital for the development of reading comprehension in primary school in Germany. We focus on mechanisms that foster reading literacy development. Therefore, we differentiate between cultural capital that refers to parental education, cultural goods (e.g., books in the household), participation in the elite culture (e.g., beaux arts), and individual reading habits. Furthermore, concerning the elite culture and reading habits, we distinguish between parents' and students' activities. We also discuss whether and how the importance of cultural capital varies between native and immigrant students.

Explanations and Previous Findings on the Importance of Cultural Capital: Social Reproduction, Cultural Mobility, and Reading Habits

Bourdieu's work (Bourdieu, 1974; Bourdieu & Passeron, 1977) is the point of departure for the concept of cultural capital. Thereby, his notion of cultural capital embraces not only educational certificates and cultural goods, but also "inculcated forms" such as abilities, skills, knowledge, and taste. Furthermore, cultural capital in an embodied state is primarily acquired in the family of the student, but as students grow older, it can also be acquired in school. Children enter the education system with different cognitive abilities and skills as well as behavior modes, which may be in part the product of class-specific socialization processes (Hart & Risley, 1992; Petrill, Deater-Deckard, Schatschneider, & Davis, 2005; Rodríguez-Brown, 2011). Bourdieu (1974; 1986; Bourdieu & Passeron, 1977) supposes that those better endowed by their families profit more from schooling and acquire new competencies much faster. In addition, the origin of the cultural capital should make a difference. Those who had the opportunity to learn from their families are designated by ease, whereas those who primarily acquired cultural capital in school are pedantic because people reveal their origin as they apply their cultural capital (Bourdieu, 1984). However, there is a lack of explicit explanation for how the transmission of cultural capital from parents to children occurs.

It is important to note that Bourdieu's (1984) approach belongs to the so-called conflict theories. According to these theories, social classes have different interests and the dominant social classes try to preserve their privileges across time and generations. In Bourdieu's version, cultural capital is crucial for securing these advantages. In general, the dominant classes impose study content, and school teachers favor children from the dominant classes because of their higher linguistic skills, specific knowledge, effort, and style. Teachers pay more attention to students from the privileged classes and give them better grades, even if they only perform as well as other students (cf. Lorenz, 2011). The function of the education system is to provide the students of the upper social classes with the highest educational degrees and students from the lower classes with lower degrees while pretending that these differences are merit based. This process of legitimation masks the intergenerational reproduction of classes, also known as social reproduction.

DiMaggio and Mohr's (1985) point of departure is Weber's (1922/1978) distinction between class and status ("Stand"). The first is defined by position and life chances in a market economy, whereas the second is defined by honor (social prestige), lifestyle, and social closure. In developed market economies, the relation between class and status is assumed to be loose, but "[t]he ability to participate in a status culture is a cultural resource that permits actors to get ahead" (DiMaggio & Mohr, 1985, p. 1235).

In other words, DiMaggio and Mohr's aim is to extend the established measure of class positions derived by occupation or educational attainment and to include indicators of status, especially of interest and participation in status culture. In this approach, parental education is an indicator of class position, whereas cultural participation is an indicator of status.

Modern societies may be characterized by affluence, democracy, mass media, consumption, and so forth. Even or just because of these conditions, status is still relevant for social positioning or achieving interests and goals. "(...) the status culture (...) retains its interactional potency for several reasons. First it has become a significant part of the formal educational system and, through that system, has been diffused, as a cultural model, throughout the class structure. Second, it is preserved through status emulation by many members of the middle class, who have adopted both the cultural tradition and the ideology that legitimates it. Third, interest in and familiarity with high culture are still related to class position, albeit imperfectly (...). Finally, high-culture activities (...) are still primarily dominated by occupants of high class positions" (DiMaggio & Mohr, 1985, pp. 1236-1237). So, what are the mechanisms relating cultural capital to students' attainment? It's "(a) increasing their opportunities for special help from teachers and other gatekeepers, (b) permitting them to develop generalized reputations as 'cultured persons', and (c) facilitating access to social milieus in which education is valued and in which information about educational opportunities is available" (DiMaggio & Mohr, 1985, p. 1240). Taken together, the mechanisms relating cultural capital to students' attainment do not highlight positive influences for academic achievement, but rather positive evaluation and recognition by significant others (Laureau & Weininger, 2003).

Comparing these different approaches, three major differences between the work by Bourdieu (Bourdieu, 1974; Bourdieu & Passeron, 1977) and the work by DiMaggio and Mohr (1985) become apparent: First, according to Bourdieu, the main function of the education system is to mask social reproduction. By contrast, DiMaggio and Mohr do not make such an assumption. Second, according to DiMaggio and Mohr, students coming from the lower or middle classes can have access to high-status culture and can profit from this access to high-status culture in terms of educational outcomes or in the labor market as well as the marriage market. In Bourdieu's theory, however, students from the lower or middle classes will not be able to change or improve their class position. Finally, DiMaggio and Mohr distinguish between parental education as a class indicator and cultural participation and interest as status indicators. We do not, however, find such a distinction in the work by Bourdieu.

Empirical Studies Relating Social Reproduction, Cultural Capital, and Educational Success

An important empirical contribution was provided by Aschaffenburg and Maas (1997), who tested rival hypotheses derived from Bourdieu's theory of social reproduction and DiMaggio's approach to cultural mobility. In their study, Aschaffenburg and Maas investigated whether and how parental cultural capital and students' participation in highbrow art impacted transitions in students' educational careers in the US. Students' cultural activities, such as performing or taking theoretically oriented lessons in music and the visual arts and taking performance classes such as in ballet and acting, were surveyed of students of different ages and thereby at different stages in the education system and by context. The context refers to activities in and outside of school. Activities in school should be accessible to all students, whereas activities outside of school should depend more strongly on the resources and initiative of the family. The four transitions under study were the beginning and termination of high school as well as the beginning and termination of college. They found that students' participation in cultural activities went hand in hand with higher probabilities of completing an educational stage and making the transition to the next educational stage. Furthermore, current activities were found to be more important for differences in the transition rates compared to earlier activities. Finally, the effects of different cultural activities were found to weaken over the educational careers of the students. Cultural activities outside the school, which may be mainly induced by the family, were found to have a stronger impact than voluntary cultural activities in school. Nevertheless, activities in school remained relevant. In addition, students' cultural activities had positive impacts on transitions in the education system even if parental capital was taken into account. Conclusively, all these findings are highly consistent with the cultural mobility model. Obviously, the assumption about social reproduction in its strictest sense – that parents' cultural capital is inculcated in children before they enter school and that advantages and disadvantages are amplified by the student's school career – does not hold. Nevertheless, in three of the four educational transitions in the study, parents' cultural participation was positively associated with transitions, giving some credit to the theory of social reproduction.

Several attempts have been made to distinguish between different types of cultural capital in order to provide further insight into the mechanisms that relate cultural capital to educational success. Thereby, the development of academic achievement has been given greater attention. Some studies, for example, have discriminated between participation in beaux arts (e.g., theater, museums) and reading behavior. The first is seen as an indication that the student belongs to some status group, which is recognized and positively valued by teachers, whereas the second is a more direct way to enhance cognitive skills (e.g., vocabulary or text comprehension). De Graaf, Dirk, De Graaf, and Kraaykamp (2000), for example, found empirical evidence from the Netherlands indicating that parental reading is relevant for educational success, more so than mere participation in the field of highbrow art. "(...) parents who read frequently not only set the norm for their children, but exhibit more human capital and therefore can enhance their offspring's linguistic and cognitive skills" (DeGraaf et al., 2000, p. 98). Comparable results were reported by Crook (1997) for Australia. In addition, Cheung and Andersen (2003) provided evidence for the long-term effects of children's reading in leisure time. They analyzed data from the British National Child Development Study (NCDS) based on a sample of children born in 1958, with surveys at ages 11, 16, 23, and 33. Children's reading behavior at age 11 was positively related to the results of a general cognitive test conducted at the same age, national school examinations at age 16, the school type attended at the secondary level, and whether the student received a university degree. Then, in a study based on German primary and secondary school students, McElvany, Becker, and Lüdtke (2009) provided evidence for a model in which different measures of social class were related to the development of reading comprehension. Major parts of these social disparities in reading comprehension were mediated by cultural resources and activities of the parents, such as visiting libraries jointly with their children or making presents of books to their children.

One of the most fine-grained studies on cultural capital was provided by Sullivan (2001). She asked students in the 11th grade about their type and amount of reading, television viewing, music listening, music playing, as well as their participation in public and cultural events in England. In the case of reading and television viewing, she coded the answers about book titles and television programs according to their cultural content. In addition, students were tested on their knowledge of famous cultural figures and on their vocabulary (Sullivan, 2001, p. 899). The students also reported on their parents' cultural activities. "These [parents'] activities include reading (and number of books in the home), newspapers taken, type of music and radio stations listened to, participation in 'formal culture', and the subjects discussed by parents in the home" (Sullivan, 2001, p. 900). Moreover, Sullivan had access to students' results in the General Certificate of Secondary Education (GCSE). Her major findings were the following: The higher the family's class position, the more culturally active were the parents and students. Furthermore, the relation between social class and students' cultural activities was mediated by parents' cultural activities. Parents' cultural activities were correlated with students' vocabulary and cultural knowledge. But if students' cultural habits were taken into account, parents' cultural activities lost their ability to predict the results of the language indicator. Sullivan's research showed that reading and watching "relatively sophisticated" television programs were positively correlated with the results in both test domains. No such positive correlation could be found for participation in cultural events and listening to classical music (including playing an instrument). Regarding the results of the GCSE, the findings were comparable at a first step, but if vocabulary and knowledge tests were taken into account in multivariate models, students' reading and television viewing did not have any contribution. In line with previous research, but relying on more detailed indicators, Sullivan (2001) concluded that the process of cultural transmission is via cognitive enhancement and not via the signaling of status membership.

Finally, studies based on data from the Programme for International Student Assessment (PISA) have provided cross-national evidence on the importance of cultural capital for educational success. In his analysis based on data from 25 Western countries, Barone (2006) found that cultural capital, which was defined as possessing culture-related goods in the family household and engaging in parent-child communication about cultural topics, was correlated with school performance in all countries. In addition, cultural goods and communication partly mediated the relation between parents' socio-economic status and students' performance. However, substantial parts of the relation between parents' socio-economic status and students' academic performance remained unexplained, thus producing the hypothesis that the applied indicators of cultural goods and activities might not be sufficient for explaining this relation. In other words, one must consider the idea that additional features related to the socio-economic situation of the parents such as ambitions and educational aspirations may also have substantial relevance.

Cultural Capital and Students' Performance in Immigrant Families

There is not much research on the importance of cultural capital for educational success in immigrant families compared to native families. Furthermore, the rare studies that have compared the importance of cultural capital of families with and without immigration backgrounds have provided evidence that is quite mixed. According to Nauck, Diefenbach, and Petri (1998), the relations between parents' cultural and economic resources as measured by the highest educational degree and the need-adjusted household income and children's secondary school attainment are much weaker in immigrant than in native families. Based on a sample of primary school students in inner London, comparable findings were reported by Strand (1999). Although the author had only a proxy indicator of the cultural and economic capital of the family (i.e., the entitlement to a free school meal), he found quite strong interactions with students' ethnic-cultural background. The social gap in students' school performance in reading, writing, and mathematics between students who received a free school meal and students who did not receive such social support was highest for non-immigrant English students. However, smaller disparities were found for students with African, Caribbean, Indian, Pakistani, and any other immigration background.

Based on data from Germany, Kristen and Granato (2007) reported weaker relations between parental education level and the child's chances of receiving a general university entrance qualification (*Abitur*) for families with Turkish origin than for native German ones. However, these results could not be confirmed when students from families hailing from Greece, Italy, Spain, Portugal, or the former Yugoslavia were considered. In her study focusing on the cultural knowledge of preschoolers aged 3 to 4, Becker (2010) provided additional insight into the role of speaking the German language for Turkish immigrant families. First, the author reported differences in the amount of cultural knowledge between children from Turkish families and natives as well as differences between children raised in families that were more or less engaged in activities such as "telling stories to child, reading books to child, (...) ever visited a zoo or circus, a library, and a museum or a theater" (Becker, 2010, p. 22). In general, children from Turkish families scored lower and children in more active families scored higher. For the Turkish students, however, Becker (2010) reported an interesting finding: A higher level of family activities went hand in hand with higher cultural knowledge scores the more often the family spoke German, the language of the receiving country. In other words, the amount of German language used by the members of immigrating families was found to moderate the effect of cultural activities on the development of cultural knowledge of the host country. The author assumed that with a higher rate of German language use in the family, the cultural content acquired by the cultural activities more and more resembled the cultural content found in families of the receiving country.

Leopold and Shavit (2013) provided a seminal contribution on the mechanisms (i.e., cognitive enhancement vs. signaling) responsible for the relation between cultural capital and school success. Therefore, they also took into account whether the cultural capital related to the country of origin of the immigrants was useful in the education system of the receiving country. The authors analyzed reading comprehension scores and grades in Hebrew and mathematics of immigrant students from the former Soviet Union and natives in Grades 4, 9, and 11 in Israel and found that "(...) immigrants and natives do not differ with regard to the effects of parental cultural capital on reading comprehension as measured by standard test scores. However, the two groups differ significantly in the effects of cultural capital on teachers' grades. The grades assigned to native students in both math and Hebrew are positively related to parents' reading behavior (as indicated by the number of books at home) and to their cultural habits, tastes, and cultural competencies, but among immigrants these relationships are much weaker or nil" (Leopold & Shavit, 2013, p. 10).

In the end, what can we conclude with regard to the role of cultural capital in educational attainment for immigrant and non-immigrant students? First, parents' human capital and certificates acquired before immigration are not always (fully) recognized in the labor market of the receiving country. This can impede economic progress and the ability to achieve higher social positions (Friedberg, 2000; Chiswick, 1978). In addition, the intergenerational transmission of cultural capital and the process of students' educational attainment might be hampered. Cultural resources are often context specific. The highbrow culture of one society might be unknown or less valued in another one; for example, the classical authors might differ in French, German, Russian, Turkish, or Vietnamese contexts. In this case, the student's knowledge of and attachment to the highbrow culture of the (parents') country of origin might not contribute to school achievement and might not serve as a signal to teachers (Leopold & Shavit, 2013).

Second, language skills can also be conceptualized as a context-specific cultural resource, which loses some of its potential in the process of immigration (Chiswick & DebBurman, 2004). Research on the importance of the use of the dominant (school) language indicates that students perform better if their family members predominantly speak the language of the receiving country at home (Kristen, 2008; Stanat & Christensen, 2006).

Third, on the other hand, there might be spillover effects of cultural capital from one language context to the other language context; for the controversial discussion of spillover effects concerning (second) language acquisition, see Cummins (2003) and Esser (2006). For some forms of cultural capital, this means that although the capital was acquired in or is related to the country of origin, it might also influence the student's educational attainment in the host country. If cultural transmission mainly takes place via habits, the language and context specificity of cultural consumption would be rather irrelevant. Parents might go on reading books written in the language used in their country of origin so that their children have an increased probability of reading too even though the children may predominantly use texts written in the language of the receiving country.

Research Questions

Although the BiKS longitudinal study provides several further possibilities, this chapter is exclusively dedicated to reading comprehension as the outcome to be explained. Furthermore, cultural capital should be more relevant for the acquisition of competencies and skills in language domains than in mathematics and science. Competencies in mathematics and science are mainly acquired in school, whereas a large proportion of the learning and practicing opportunities in the language domain is provided by the family. In this chapter, we focus on three major research questions:

1) What is the contribution of different forms of cultural capital on students' reading comprehension?
2) Does the impact of cultural capital on reading comprehension differ between students from native and immigrant families?
3) Do we find that the evidence favors the social reproduction theory or the social mobility model? According to the social reproduction theory, a child's cultural capital and school performance is a direct function of the parents' cultural capital (formal education, cultural activities, possession of cultural goods, etc.), whereas the cultural mobility model gives special credit to the child's activities. In contrast to DiMaggio and colleagues, who emphasized only the signaling effect of cultural activities, we further extended the social mobility model to the effect of cognitive stimulation on students' cognitive development.

In order to provide answers to these three questions, we distinguished between parental education, number of books in the household as cultural goods (reproduction), and the children's use of libraries (mobility), the children's and parents' highbrow activities as well as the amount of reading, and how much the German language is used in families with immigration backgrounds.

Method

Sample

All analyses refer to data collected within the framework of the Bamberg BiKS-8-14[1] longitudinal study. The interdisciplinary BiKS research group, founded in 2005, consists of researchers from disciplines such as education, psychology, and sociology (cf. Lorenz, Schmitt, Lehrl, Mudiappa & Roßbach, chapter 2, this volume). In this chapter, we used data from the second cohort, which traced the development of students from the third grade up to the ninth grade (cf. von Maurice et al., 2007). In total, data from $N = 2{,}395$ primary school students attending 155 classes at 82 different schools were available. In elementary school, students were tested three times. The first measurement point took place at the beginning of the second term of Grade 3. Consecutive measurement points took place in the middle of the first term of Grade 4 and finally at the end of the second term of Grade 4. After the transition into secondary school, data collection took place annually at the end of the academic year. Students were tested with a broad battery of competence measures. In addition, student data collected through standardized questionnaires were available. The students' parents participated in a computer-assisted telephone interview (CATI). And finally, a questionnaire for the students' teachers comprising questions about the school class composition, teaching methods, and the educational background of the teacher as well as questions about individual children participating in the study was administered.

The current chapter presents data from the first and third measurement point, when the students attended the third and fourth grades, respectively. Cases with unit nonresponse, which includes both students who had not been tested and parents who had not provided an interview at one or both testing points, were excluded from all analyses ($n = 785$; 32.8%). Further, $n = 136$ (5.7%) cases were excluded due to item nonresponse, resulting in a final sample of $n = 1{,}474$ students and their parents used in our analyses. Parents respectively students remaining in the sample differed in some characteristics from those being excluded: For example, parents remaining in the

---

[1] BiKS is the acronym for the German title "Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vor- und Schulalter," which means "Educational Processes, Competence Development, and Selection Decisions in Preschool and School Age" in English.

sample were better educated (13.7 vs. 12.9 years) and had less often an immigration background (17.4 vs. 36.1%). Students remaining in the sample performed better in the first reading test (48.1 vs. 51.3).

The average age of the students in the analyzed sample was 9.2 years in Grade 3 and 10.3 years in Grade 4. Seven hundred sixty-eight (52.1%) students were male and 706 (47.9%) students were female.

Measures

The dependent variable was *reading comprehension* at the end of the fourth grade (measured at the third measurement point). Because we focused on changes in reading comprehension, we also took into account reading comprehension in the middle of the third grade (measured at the first measurement point). At the first measurement point, reading comprehension was measured by a sample of 13 short texts with 20 multiple-choice items from the subscale "text comprehension" of the "Ein Leseverständnistest für Erst- bis Sechstklässler" (ELFE 1-6; Lenhard & Schneider, 2005). At the third measurement point, the ELFE subscale "text comprehension" was lengthened by adding three new texts with six multiple-choice items developed by the BiKS research group. This test elongation was necessary in order to avoid ceiling effects. For the reading comprehension test, the students had to read a given text, search the relevant information, and generate inferences from the text to answer the given items. Test time was limited to 7 min for the entire reading comprehension test. The item difficulty parameters were estimated within an IRT framework assuming a 1-parameter Rasch model with a Gaussian population distribution. In a first step, item difficulty parameters were estimated for the 26 reading comprehension items used at the third measurement point. Subsequently, the item difficulty parameters of the 20 reading comprehension items used at the first measurement point were fixed to guarantee a common metric. The individual student's ability was estimated by Weighted Likelihood Estimates (WLEs) using the ConQuest software package (Wu, Adams, Wilson, & Haldane, 2007). WLE scores were subsequently T-standardized ($M = 50$, $SD = 10$) based on the first measurement occasion. The internal consistencies (Cronbach's α) of the measures were satisfactory for all time points ($\alpha_{\text{time 1}} = .88$, $\alpha_{\text{time 2}} = .87$, and $\alpha_{\text{time 3}} = .89$).

In order to take parents' cultural resources into account, the highest qualification they achieved was used, measured in *years of education*. This scale covers the typical institutional time spent in school, vocational training, and tertiary education for achieving a certain qualification and ranges from 7 years for no formal certificate up to 18 years for a university degree (cf. Helberger, 1988).

As an indicator of cultural possessions in the family, we relied on the *number of books in the household*. Parents reported possessing *no* (codes as 0), *less than 11* (1), *11 to 50* (2), *51 to 100* (3), *101 to 250* (4), *251 to 500* (5), or *more than 500 books* (6).

Parents were asked if the child reads for pleasure. The possible answers were yes, every day (coded as 3), yes, several times a week (2), less often (1), or hardly ever or never (0).

Parents provided information about the *child's cultural activities* at the first measurement point (third grade). They indicated how often they attended the following together with their child during the last year: (a) museums, (b) libraries, (c) kids' concerts, (d) kids' theaters, (e) zoos or wildlife parks. The possible answers were *at least once a week*, *at least once a month*, *several times a year*, *less often*, and *never*. Although exploratory factor analysis yielded only one factor, only the items for museums, kids' concerts, and kids' theaters showed high factor loadings, whereas the items for libraries and zoos had relatively low loadings. Consequently, the three items measuring the *child's highbrow culture* were summed to form one scale (Cronbach's $\alpha = .60$). The scale ranged from 0, indicating no activity at all, to 4, indicating – at least hypothetically – weekly activities in all three domains. The *visits to libraries* item was used as a single-item indicator. The scale ranged again from 0, indicating no activity at all, to 4, indicating weekly library visits. Library visits might be an alternative or a supplement to possessing books and therefore served as an appropriate indicator of cultural mobility. The zoo item was disregarded because it was not linked to the concept of cultural capital.

The *parents' cultural activities* were measured at the third measurement point (end of the fourth grade). The introduction of the measures on cultural participation mentioned whether the interviewee attended cultural events alone or together with his or her child. In the subsequent questions, the parent was asked whether he or she had visited the following events or sites during the last year: (a) an art or historic museum,

(b) an exhibition, (c) a cabaret, theater, ballet, or an opera performance, (d) a classical concert (as well as other concerts and courses unrelated to job/career). For each affirmative answer, he/she reported subsequently how often he/she had attended such places/shows in the last year on an open-ended scale. As the distributions of the answers on these items were highly right skewed, we transformed the scale by first adding 1 to every answer and then taking the natural logarithm. People who did not attend cultural activities at all still received a value 0 after this transformation (as $\ln(1) = 0$). The four items on the *parents' highbrow culture* were summed into one scale (Cronbach's α = .64). We should mention that the parents' and child's cultural activities might overlap to some degree. The measure of the child's cultural participation was clearly defined (e.g., child's theater), whereas the introduction of the item block on the parents' cultural participation also mentioned the child. However, the items were targeted to adults to a higher degree (exhibition, opera).

The parent also reported how many hours he/she had *read newspapers* or *books* during the last month. The answers to both questions were only weakly correlated and were therefore used separately in the analyses. Because the reports on hours of reading newspapers or books during the last month were right skewed as was also the case for the number of highbrow activities, we transformed and logarithmized the answers as already described above. We assumed that both the parents' visits to highbrow events and their reading behavior would remain stable over time and would not be influenced by the child's progress in text comprehension and that it would therefore be justifiable to use them as predictors even though they were surveyed at the third measurement point.

Families with at least one parent born abroad were considered to have an *immigration background*. In these cases, we also indicated whether the family reported speaking with the child (a) always in German, (b) mostly in German, (c) in German and another language to the same degree, or (d) mostly in another language/other languages. Each of these categories was coded using binary variables.

The parents' education and cultural activities and habits could also be indicators of the family's economic situation. In order to avoid a confounding influence, we focused on the monthly disposable household income including state transfers. Because income is a sensitive question with a large proportion of item nonresponses and therefore also might have a reduced reliability, we took the average of all valid pieces of income information collected at the first three survey measurement points. However, in about every tenth case, there was still no income information. Therefore, we applied a regression-based single imputation to fill the gaps. Income was need-adjusted by the square root of the number of persons in a household; due to the positive skew of the distribution, we used the logarithms of the income values.

Student's *gender* was dummy coded 1 for male and 0 for female students.

Finally, we also controlled for *general cognitive abilities* measured at the first measurement point. Students' general cognitive abilities were assessed with a set of 15 items from the matrices subtest of the Culture Fair Intelligence Test (CFT; German version: Weiß, 2006). This test measures the ability to recognize and solve problems of figural relations and of formal figural reasoning with different levels of complexity within a time limit of 3 min. General cognitive abilities have a strong heritable component (Bouchard & McGue, 1981; Plomin & Spinath, 2004), but are not independent of influences from the school (Becker, Lüdtke, Trautwein, Köller, & Baumert, 2012). When controlling for students' general cognitive abilities, we tried to avoid biased parameter estimations due to genetic covariation between students' cognitive abilities and parents' background.

Statistical Methods

As the data consisted of students (i) in school classes (j), we estimated multilevel linear regressions with a random intercept. The dependent variable $Y_{ij,t+1}$ is the reading comprehension of each single student (i) at the end of the fourth grade measured at the third measurement point. As we were interested in reading progress, we controlled for reading comprehension $Y_{ij,t}$ in the third grade measured at the first measurement point. Further covariates were all measured at the individual level. They refer to the student or his/her family. All unobserved characteristics imposing the same influence on test results at both points of measurement were cancelled out by controlling for the results of the first tests. This procedure reduces biased estimations for students' and family's activities and characteristics due to unobserved heterogeneity.

As there could be substantial differences in reading comprehension in Grade 4 between school classes due to factors such as class composition, quality of instruction, teacher characteristics, and so on, we estimated random intercept models. As variation between school classes was not in our research focus, we did not add any covariates to the second level (j). However, allowing for variation between classes reduces the risk of biased estimations for coefficients and their standard errors at the individual level (cf. Hox, 2002; Nezlek, Schröder-Abé, & Schütz, 2006).

Results

First, some descriptive statistics and correlations are presented. Subsequently, the results of the multivariate analysis are reported.

Descriptive Findings on Reading Comprehension in Grades 3 and 4 and Correlations between Different Indicators of Cultural Capital

In a first step, a short overview of the characteristics of the two subsamples of students (immigrant and non-immigrant students) is provided (see Table 1). Average reading comprehension scores according to the main characteristics at both points of measurement are presented. The values of the variables for parents' education, children's and parents' highbrow visits, parents' reading behavior, and household income were aggregated for this overview.

Table 1. Sample description: average reading comprehension by main characteristics for students without and with an immigration background at the first (Grade 3) and third (Grade 4) measurement point (balanced panel)

Subsample	Native students			Immigrant students		
	M (SD)	*M (SD)*	*n*	*M (SD)*	*M (SD)*	*n*
Grade	3	4		3	4	
total	51.85 (9.97)	63.78 (12.36)	1,218	48.78 (9.18)	61.18 (11.45)	256
parents' years of education						
7 to 10	48.69 (7.60)	57.21 (9.90)	15	47.30 (7.61)	57.21 (12.01)	22
>10 & <13	50.16 (9.32)	61.13 (11.32)	629	46.78 (8.59)	58.74 (9.49)	101
13 to 16	52.72 (9.39)	65.20 (12.31)	273	48.59 (8.92)	60.32 (10.89)	60
>16 (i.d.18)	54.76 (11.11)	68.34 (13.02)	301	52.15 (9.8)	66.45 (12.59)	73
# of books at home						
1 to 10	42.15 (6.82)	51.10 (15.00)	4	44.25 (4.21)	51.45 (8.00)	7
11 to 50	47.53 (7.88)	58.00 (11.02)	84	43.41 (7.93)	56.02 (8.88)	45
51 to 100	49.11 (8.58)	60.86 (11.77)	172	47.62 (7.75)	58.07 (10.44)	47
101 to 250	51.43 (9.24)	62.82 (11.32)	378	49.65 (8.62)	63.08 (10.38)	75
251 to 500	52.73 (10.81)	65.64 (12.56)	332	51.79 (10.13)	62.82 (11.67)	46
> 500	54.85 (10.36)	66.90 (13.04)	248	52.21 (9.76)	67.51 (13.45)	36
child reads for joy						
hardly ever/never	43.83 (7.01)	54.35 (9.37)	130	44.38 (7.03)	53.69 (8.29)	37
less often	47.51 (7.97)	58.96 (10.98)	207	45.49 (8.37)	58.44 (8.71)	50
several times a week	51.34 (8.81)	63.33 (10.48)	322	49.12 (9.18)	60.76 (9.80)	70
every day	55.62 (10.06)	68.01 (12.55)	559	51.84 (9.19)	65.65 (12.85)	99
highbrow part. (child)						
0	49.20 (8.05)	59.63 (11.17)	113	48.81 (9.33)	58.49 (8.57)	47
>0 to 1	51.37 (10.01)	63.01 (12.07)	666	47.55 (8.92)	59.98 (11.17)	132
>1 to 2	53.30 (10.17)	65.96 (12.45)	422	50.31 (9.13)	64.23 (12.12)	66
>2	52.23 (10.66)	67.13 (17.91)	17	54.16 (9.87)	68.72 (15.52)	11
child's library use						
never	50.58 (9.63)	61.88 (12.03)	395	47.61 (8.78)	59.21 (9.75)	87
less often	50.45 (9.13)	62.83 (11.58)	143	49.15 (9.44)	63.22 (13.27)	27
several times a year	51.62 (9.46)	63.54 (12.77)	213	48.91 (10.53)	59.73 (12.66)	47
at least once a month	52.93 (10.80)	64.93 (12.21)	334	49.84 (8.76)	64.10 (11.34)	69
at least once a weak	54.81 (9.69)	67.93 (12.70)	133	49.26 (8.97)	60.51 (11.68)	26
highbrow part. (parent) (p.a.)						
0	49.36 (8.14)	60.58 (11.31)	189	47.10 (8.13)	56.58 (9.44)	66
>0 to 1	51.35 (9.74)	62.74 (11.88)	687	48.55 (9.75)	61.82 (11.39)	128
>1 to 2	54.46 (10.87)	67.44 (13.26)	238	49.55 (8.25)	63.27 (11.46)	41
>2	53.76 (10.91)	68.04 (12.26)	104	53.91 (9.00)	67.63 (13.10)	21
reading newspaper (parent) (hours/month)						
0	51.20 (10.08)	62.93 (10.86)	90	47.85 (7.01)	58.52 (9.88)	30
>0 to 7.5	51.56 (9.63)	63.69 (11.98)	400	50.52 (9.64)	62.44 (11.24)	78
>7.5 to 15	52.42 (10.24)	64.26 (12.81)	556	48.12 (9.38)	61.43 (11.6)	98
>15	51.02 (9.80)	62.87 (12.50)	172	47.90 (9.07)	60.30 (12.36)	50
reading books (parent) (hours/month)						
0	50.19 (8.89)	61.30 (11.73)	306	47.40 (8.12)	60.04 (11.87)	81
>0 to 7.5	52.02 (10.4)	64.56 (11.89)	210	49.01 (10.46)	61.51 (11.47)	40
>7.5 to 15	53.20 (10.63)	64.97 (13.38)	347	50.89 (9.68)	63.44 (11.85)	62
>15	51.87 (9.78)	64.28 (11.85)	355	48.38 (8.97)	60.33 (10.54)	73

immigration background + language use						
mostly non-German				48.09 (8.70)	60.08 (12.48)	30
German as often as others				49.12 (9.99)	60.61 (11.05)	47
mostly German				47.21 (9.32)	60.54 (12.38)	82
only German				50.15 (8.7)	62.32 (10.55)	97
natives	51.85 (9.97)	63.78 (12.36)	1,218			
household income						
1 (lowest quintile)	49.19 (8.63)	61.02 (11.16)	241	46.94 (8.29)	58.48 (10.29)	82
2	50.80 (10.13)	61.48 (12.23)	246	46.95 (11.13)	59.74 (10.30)	47
3	52.76 (9.39)	64.99 (11.53)	250	48.46 (7.57)	62.22 (11.46)	52
4	53.20 (10.12)	65.34 (12.55)	228	50.54 (7.95)	60.74 (11.90)	34
5 (highest quintile)	53.31 (10.86)	66.03 (13.37)	253	53.50 (9.69)	67.26 (12.60)	41
gender						
female	52.61 (9.85)	65.85 (12.14)	585	49.44 (9.85)	63.29 (11.78)	121
male	51.15 (10.04)	61.86 (12.26)	633	48.19 (8.53)	59.28 (10.85)	135

Source: BiKS 8-14, measurement points 1 to 3, our own calculations.

In total, 256 students (17.4%) had immigration backgrounds, and 1,218 students (82.6%) in the sample were natives. On average, both groups made substantial progress in reading comprehension over time, but students from immigrant families scored lower on reading comprehension in comparison to native students at both measurement occasions. For parental education and the number of books in the household, the results provided a clear picture: The higher the parents' formal qualifications or the more books available in the home, the higher the average reading comprehension scores of students from both groups and at both measurement points. A similar pattern was observed for the children's amount of time spent reading and children's attendance of highbrow performances. A different pattern, however, was found concerning the frequency of joint library visits: Whereas mean reading comprehension scores steadily increased with the frequency of joint library visits for native students, such a clear pattern was not found for students with immigration backgrounds.

Regarding parents' activities, a trend toward increasing reading comprehension scores with increasing parental highbrow cultural activities was found for both immigrant and non-immigrant students. However, the relation between the amount of time parents spent reading newspapers or books and students' reading comprehension was nonlinear. In most cases, children had the highest results if parents read newspapers or books 7.5 to 15 hours a month (equivalent to 15 to 30 min per day). If parents indicated reading more or less, children fared less well in most cases. For students with at least one parent born abroad, we also display the average reading comprehension and the proportion of German language use in the family. There was no clear pattern for this family indicator. The higher the disposable household income, the higher the average reading comprehension scores within both subsamples (immigrant and non-immigrant students). And finally, the average reading comprehension scores differed between boys and girls: Girls outperformed boys independent of immigration status.

The correlations between the different indicators of cultural capital are presented in Table 2. The correlations were calculated separately for each immigration status. Correlations for natives are below the diagonal, and correlations for children of immigrants are above the diagonal. The strongest correlations were found between parental education and the number of books in the household (.53 and .57) as well as between the child's and the parents' visits of highbrow cultural events (.57 and .54). Note that in the case of highbrow culture, the constructs might not be distinct, i.e., they might overlap (cf. discussion in the data and method sections). Furthermore, in both native and immigrant families, there were additional considerable correlations between the parents' education level, the number of books in the household, the students' and parents' cultural participation, and the parents' amount of time reading books. All other correlations were below .30.

Table 2. Correlations between different types of cultural resources and activities for natives (below diagonal) and families with an immigration background (above diagonal)

			1	2	3	4	5	6	7	8
1		parents' years of education	1	.57	.10	.32	.20	.45	.14	.33
2		# of books at home	.53	1	.25	.31	.09	.44	.15	.32
3	child	reads	.19	.19	1	.28	.23	.27	-.01	.08
4	child	highbrow participation	.36	.37	.18	1	.29	.54	.14	.18
5	child	library	.12	.11	.16	.22	1	.24	.09	.18
6	parent	highbrow participation (ln)	.39	.44	.18	.57	.18	1	.25	.22
7	parent	newspapers (ln)	.09	.10	.02	.07	.06	.12	1	.11
8	parent	books (ln)	.21	.32	.10	.19	.14	.21	.14	1

Source: BiKS 8-14, measurement points 1 and 3, natives: n = 1,218, immigrants: n = 256; our own calculations.

Taken together, the large number of small correlations between the different indicators suggests that the indicators capture different aspects of cultural capital in the family and that parents as well as children show substantial differences in their amount of cultural capital.

Multivariate Analysis on the Importance of Cultural Capital for Progress in Reading Comprehension

This section contains the results of the multivariate models predicting students' reading comprehension at the end of Grade 4. First, we estimated all models separately for students with and without immigration backgrounds (see Table 3). In a first step, we included only the variables parents' education and number of books in the household, which are common indicators of cultural capital in educational research. In a second step, we introduced the control variables disposable household income, general cognitive ability, and gender. In a third step, we controlled for previous reading comprehension measured in the third grade. This means we shifted from purely cross-sectional to value-added models, controlling for unobserved heterogeneity to a much greater extent. Finally, the last column displays a model restricted to students with at least one parent born abroad. This model was extended by the share of German language use in the family (Model M4i).

Table 3. The importance of parents' education and number of books for reading comprehension at the end of the fourth grade in native and immigrant families; results of random-intercept models

	Natives M1n *b/(SE)*	Immigr. M1i *b/(SE)*	Natives M2n *b/(SE)*	Immigr. M2i *b/(SE)*	Natives M3n *b/(SE)*	Immigr. M3i *b/(SE)*	Immigr. M4i *b/(SE)*
parents' years of education	0.91**	0.55*	0.85**	0.40	0.52**	0.28	0.25
	(0.14)	(0.25)	(0.14)	(0.25)	(0.10)	(0.19)	(0.19)
# of books at home	1.10**	2.09**	0.88**	1.55*	0.07	0.45	0.45
	(0.34)	(0.60)	(0.33)	(0.61)	(0.24)	(0.47)	(0.48)
household income (ln)			0.27	1.82	-0.53	0.59	0.64
			(0.90)	(1.53)	(0.64)	(1.16)	(1.22)
cognitive ability			0.80**	1.22**	0.12	0.41+	0.41+
			(0.14)	(0.29)	(0.10)	(0.23)	(0.23)
boy (girl)			-3.82**	-3.32**	-2.71**	-2.78**	-2.83**
			(0.66)	(1.29)	(0.47)	(0.98)	(0.99)
reading comprehension (in the third grade)					0.86**	0.80**	0.80**
					(0.02)	(0.06)	(0.06)
language use (only German)							
mostly German							0.76
							(1.22)
~50/50							-0.14
							(1.44)
mostly non-German							0.24
							(1.71)
constant	46.49**	45.66**	41.40**	27.63*	16.58**	10.42	9.83
	(1.72)	(2.91)	(6.39)	(11.00)	(4.61)	(8.43)	(9.03)
variance							
class level	2.85	0.00	2.13	3.47	4.04	0.00	0.00
individual level	134.61	113.13	127.73	96.76	64.14	58.31	58.09
rho	0.02	0.00	0.02	0.03	0.06	0.00	0.00

Source: BiKS 8-14, measurement points 1 to 3, our own calculations.
Case numbers: 1,218 native students out of 149 school classes, 256 students of immigrant families out of 113 school classes.
Notes. Reference categories in italics; significance levels: $^{+}$ $p \leq .10$, * $p \leq .05$, ** $p \leq .01$.

In the first models, M1n and M1i, positive and significant coefficients were estimated for parents' education as well as the number of books in the household. As in the descriptive statistics depicted in Table 1, higher formal education and more books in the household were related to higher test results for both immigrant and native students. Including the control variables in the second set of models, M2n and M2i, led to a reduction in the size of the coefficients for parental education and number of books in the household. In the case of students with a least one parent born abroad, the coefficient for parental education failed to reach significance.[2] In the third set of models, M3n and M3i, the analysis shifted to a focus on the differential progress in reading comprehension as students' reading comprehension in Grade 3 was added as a covariate. In this model, parents' education level remained a significant predictor of reading comprehension development within the subsample of native students but not within the subsample of students with an immigration background. In addition, the number of books in the household did not make any difference in the growth of reading comprehension in both subsamples. Regarding language use in immigrant families, model M4i did not show different progress in reading comprehension in relation to the amount of German language use in the family.

Models M3n and M3i served as references for the next set of analyses. Each model was expanded by only one indicator. We began with the indicators of the mobility approach, namely, children's activities, and then added indicators of the social reproduction approach, parental activities (see Table 4). All effects were estimated under the control of parents' education, number of books, household income, students' general cognitive abilities, previous reading achievement, and gender. In both subsamples, there was a significant positive relation between students' amount of reading and the development of reading comprehension. Regarding students' highbrow cultural activities, the coefficients in both subsamples were positive (more activities led to higher growth), although only the coefficient estimated for the immigrant subsample was significant. The coefficient for students' highbrow cultural activities in the native subsample did not reach statistical significance. Students' frequency of library visits was not related to the development of reading comprehension. For the parents' frequency of visiting highbrow events, positive effects of the development of reading comprehension were estimated. In the native subsample, the effect was significant at the 10% level and in the immigrant subsample, at the 5% level. Therefore, the amount of parents' cultural activities was positively linked to students' development of reading comprehension. However, the coefficient estimated for the immigrant subsample was nearly three times as large as the estimated coefficient for the native subsample. Parental reading behavior was not linked to students' growth in reading comprehension as the

[2] Interestingly, the coefficients for household income were not significant. Financial resources seem to be unrelated to reading comprehension.

estimated coefficients did not reach significance at the 5% level. In the immigrant subsample, one negative coefficient for parents' amount of book reading was found. This coefficient was significant at the 10% level.

Table 4. Effects of cultural participation and activities on progress in reading achievement in native and immigrant families – enlargement of models M3n and M3i by one variable[1]

	child			parent		
	reads (0-3)	highbrow part. (0-3)	uses library (0-4)	highbrow part. (ln)	reads news-paper (ln)	reads books (ln)
Natives						
Coeff.	0.82**	0.49	0.23	1.23⁺	-0.02	0.01
SE	(0.25)	(0.45)	(0.17)	(0.65)	(0.03)	(0.02)
Immigrants						
Coeff.	1.33**	1.94**	0.14	3.46*	0.01	-0.05⁺
SE	(0.49)	(0.73)	(0.36)	(1.36)	(0.05)	(0.03)

[1] All models include variables on parents' education, number of books, household income, students' general cognitive abilities, previous reading achievement, and gender, see Table 3.
Source: BiKS 8-14, measurement points 1 to 3, our own calculations.
Case numbers: 1,218 native students out of 149 school classes, 256 students of immigrant families out of 113 school classes.
Notes: Reference categories in italics; significance levels: $^{+}$ $p \leq .10$, * $p \leq .05$, ** $p \leq 0.01$.

Taken together, both students' frequency of library visits and parents' reading behavior did not promote progress in reading comprehension. The student's own reading behavior, however, positively influenced growth in reading comprehension. Students' and parents' attendance of highbrow events also seemed to have an impact on reading progress; this effect was especially pronounced for students raised in immigrant families. Remember, according to Table 3, parents' formal qualifications seemed to have lower or even no influence in immigrant families. These findings strongly suggest that the importance of cultural capital differs between native and immigrant families. However, the two subsamples differed considerably in sample size, and the standard errors of the point estimates were only considered superficially. In addition, a few effects might be spurious and might disappear after controlling for other forms of cultural capital.

Finally, models comprising both subsamples were estimated. Immigration status was included in the models as a predictor variable. The models depicted in Table 5 included all variables already used in models M3n and M3i (see Table 3) plus a binary variable indicating students' immigration status. Furthermore, variables with significant coefficients in Table 4 were added to the model. These variables consisted of students' time spent reading as well as the students' and parents' amount of participation in highbrow events. For all of these variables, main effects were estimated and displayed in Model 5 (Table 5). The next three models included an additional interaction term between immigration background and parental education (Model 6) as well as the students' or the parents' amount of participation in highbrow events (Models 7 and 8, respectively). In the last model (Model 9), all three interaction terms were included simultaneously.

Table 5. The importance of cultural capital and immigration background for reading comprehension at the end of the fourth grade – value-added models with random intercepts

	M 5 *b/(SE)*	M 6 *b/(SE)*	M 7 *b/(SE)*	M 8 *b/(SE)*	M 9 *b/(SE)*
Immigration background (*native*)	0.10 (0.57)	1.12 (2.45)	-0.95 (0.93)	-0.60 (0.85)	1.62 (2.50)
parents' years of education	0.41** (0.09)	0.43** (0.10)	0.41** (0.09)	0.41** (0.09)	0.46** (0.10)
Interaction term with immig.		-0.07 (0.17)			-0.23 (0.20)
# of books at home	-0.11 (0.22)	-0.11 (0.22)	-0.11 (0.22)	-0.11 (0.22)	-0.10 (0.22)
child reads for joy	0.86** (0.23)	0.85** (0.23)	0.84** (0.23)	0.85** (0.23)	0.83** (0.23)
highbrow part. (child)	0.35 (0.43)	0.35 (0.43)	0.04 (0.48)	0.33 (0.43)	0.04 (0.50)
Interaction term with immig.			1.19 (0.82)		1.11 (0.98)
highbrow part. (parent)	1.16+ (0.66)	1.17+ (0.66)	1.22+ (0.66)	0.92 (0.69)	1.02 (0.72)
Interaction term with immig.				1.52 (1.35)	1.30 (1.70)
household income (ln)	-0.37 (0.56)	-0.38 (0.56)	-0.35 (0.56)	-0.37 (0.56)	-0.38 (0.56)
cognitive ability	0.16+ (0.09)	0.17+ (0.09)	0.16+ (0.09)	0.16+ (0.09)	0.16+ (0.09)
boy (*girl*)	-2.43** (0.43)	-2.43** (0.43)	-2.45** (0.43)	-2.44** (0.43)	-2.46** (0.43)
reading comprehension (at 3rd grade)	0.82** (0.02)	0.82** (0.02)	0.82** (0.02)	0.82** (0.02)	0.82** (0.02)
constant	16.40** (4.08)	16.25** (4.10)	16.52** (4.08)	16.63** (4.09)	16.23** (4.10)
variance					
class level	1.22	1.08	1.13	1.19	1.12
individual level	62.88	62.88	62.88	62.89	62.84
rho	0.02	0.02	0.02	0.02	0.02

Source: BiKS 8-14, measurement points 1 to 3, our own calculations.
Case numbers: 1,474 students out of 153 classes.
Notes. Reference categories in italics; significance levels: $^{+}$ $p \leq .10$, * $p \leq .05$, ** $p \leq .01$.

Model 5 did not show any overall differences in the growth of reading comprehension between students of native and immigrant families. There was again a positive highly significant effect of the child's amount of time spent reading on the progress in reading comprehension. Regarding highbrow activities, the main effect of parents' activities was significant at the 10% level, whereas the main effect of students' highbrow activities failed to reach significance. However, both indicators were highly correlated (see Table 2). Therefore, if we included only one of these two indicators, parents' activities were significant at the 5% level and children's activities at the 10% level (results are not shown in the table). Consequently, it seems that parents' highbrow cultural activities have a stronger impact on students' reading comprehension than students' own cultural engagement.

Regarding the interaction effects in Models 6 through 9, all of them pointed in the direction suggested by the previously estimated models, but none of the interaction effects was statistically significant. In addition, the comparison of the remaining unexplained variance on the individual level in Model 9 with the individual variance in Model 5 revealed that the interaction terms did not reduce the unexplained variance at the student level. Therefore, the spare Model 5 should be preferred to Model 9, which contained three additional interaction terms. Consequently, the results of the joint analytic model did not provide support for immigrant-specific differences in the importance of cultural capital for progress in reading comprehension at the end of primary school.

Conclusions

At the end of the theoretical introduction on the importance of cultural capital, we posed three main research questions. In the following section, we will discuss every research question separately with regard to the presented results.

The first research question of this study concerned the contributions of different forms of cultural capital to the student's reading comprehension. In order to gain insight into this topic, we decided to investigate progress in reading comprehension instead of merely analyzing reading comprehension at a single point in time. The focus on explaining differences in progress reduces the threat of biased estimations and the problem of reversed causation. For example, students who like reading a lot might do so because they are excellent readers and reading is easy for them. The advantage of value-added models can also have some drawbacks as previous positive influences on the status achieved at the first measurement point cannot be discovered. Consequently, results are conservative (i.e., we might have underestimated the influence of relevant factors). In addition, our empirical analyses still relied on nonexperimental data.

Unobserved factors with time-varying influences might be correlated with our variables and therefore might bias the estimated effects.

Our empirical analyses on the development of reading comprehension from the middle of the third grade to the end of the fourth grade indicate that students perform better over time the higher their parents' educational level is and the more time the students themselves spend reading. There is also evidence that participation in highbrow culture promotes growth in reading comprehension, especially the parents' participation in such activities. No effects could be found for the number of books in the household or children's use of libraries. The number of books is an indicator of the opportunity structure. Library visits are also an indicator of the opportunity structure, but might also be an indicator of interest in reading. In addition, parental reading time was not related to the child's competence gains, even if we did not control for the child's own reading time.

How should these findings be interpreted? First, inequalities in reading comprehension increased between children raised in families with lower and higher educational backgrounds during the last year of primary school. Second, this widening gap could not be fully explained by the reading habits or cultural activities of the students or their parents. This means that relevant indicators for explaining the widening gap were missing from our analyses. Third, the student's amount of reading had a positive impact on progress in reading comprehension, but the parents' amount of reading did not. Furthermore, the student's and parents' amount of time spent reading were not substantially correlated with each other. These findings call for a cautious view of simple models that assume that parental reading behavior serves as a role model and is simply reproduced by the students. The findings also raise concerns about the fact that parental reading as such produces a more stimulating literacy environment for the child (e.g., different vocabulary, more complex grammar). However, the available indicators differentiated only between reading newspapers and reading books. Nevertheless, there was no indication of the quality of this reading material, limiting the explanatory power of this finding. This leads directly to the fourth point: the attendance of highbrow cultural events (e.g., theater, classical concerts, etc.). Parents' and students' frequency of engagement in these activities were strongly correlated with each other. This seems quite plausible as the students under investigation were of primary school age, and these activities should be highly influenced by parents' contributions to students' leisure time activities. In addition, the parents' participation in beaux arts exerted a stronger influence on the student's progress in reading comprehension than on the student's own participation in beaux arts. In contrast to the indicators of the parents' reading behavior, the participation in highbrow culture was more clearly related to cognitively demanding activities. These activities seem to enhance competencies in the language domain. However, such activities might also be based on some third variables such as higher parental skills and cognitive capacities, which could also lead to a more stimulating home environment for the student. Therefore, we should be careful about making causal interpretations of these finding.

The second research question referred to differences in the impact of cultural capital on reading comprehension between native and immigrant students. With regard to the existent literature (e.g., Nauck, Diefenbach, & Petri, 1998), we expected stronger relations between measures of cultural capital and academic achievement for native students than for students with immigration backgrounds. This expectation was partially confirmed. Whereas parental education background was significantly related to the development of reading comprehension in the subsample of native students, no such relation was found in the subsample of students with an immigration background (Models M3n and M3i, Table 3). Therefore, it seems that cultural capital in terms of educational level acquired in a foreign country is not as easily transferred to the next generation as the same type of cultural capital acquired in the host country by native parents. However, in a joint model, the interaction term of the educational background of the parents and immigration status did not reach significance (Model M9, Table 5). Therefore, the result of different influences of the educational background of the parents with and without an immigration background on the development of reading comprehension should be interpreted with caution. With regard to cultural activities, the opposite seems true: Participation in highbrow cultural activities was more highly related to reading comprehension for students from immigrant families than for students from native families (cf. Table 4). The tested interaction effects in the joint model, however, also did not confirm these findings from the separate analyses for students with and without an immigration background.

Therefore, although only of preliminary status, we might conclude that some forms of cultural capital, especially more distal aspects such as the parents' educational background, are of higher importance for students from native families than for students with an immigration background. Behavioral aspects such as the participation in highbrow cultural activities within the host country, however, seem at least equally influential for the educational attainment of both groups – immigrant students as well as native students. This is consistent with our expectations: Cultural capital in terms of parents' level of education that was acquired in a foreign country is often less directly transferable into students' educational success in another country. Participation in highbrow cultural activities, however, at least as these activities were measured in the BiKS-8-14 study, takes place in the host county and therefore can be more directly converted into the educational success of the students. Finally, a specific feature of the immigrant families in this study was that the majority indicated that they do not use the receiving country's language (German) at home. However, we might consider the use of German language itself as a specific aspect of cultural capital. According to our analysis, there was no difference in the progress in reading comprehension with regard to the amount of German language use in the family (Model M4i, Table 3). This was contrary to our expectations, as the language spoken in the family has been shown to be a relevant predictor of academic achievement, including students' reading competence level (Müller & Stanat, 2006; Stanat & Edele, 2011).

Finally, with the third research question, we asked whether results from the BiKS-8-14 longitudinal study were consistent with the model of social reproduction (cf. Bourdieu, 1974; Bourdieu & Passeron, 1977) or whether our results could provide support for a model of social mobility (cf. DiMaggio, 1982; DiMaggio & Mohr, 1985). The findings on activities of highbrow culture may be interpreted in favor of social reproduction theory in the tradition of Bourdieu. The parents' and children's highbrow cultural activities were highly correlated and the parents' highbrow cultural activities imposed a stronger influence on the progress in reading comprehension. The impact of the parents' formal qualifications on progress in reading comprehension could also be credited toward the theory of social reproduction. However, the strong influence of reading habits, independent of the parents' cultural activities and educational level, counts toward cultural mobility in a broader sense. In a strict sense, in the version offered by DiMaggio and Mohr, cultural mobility is mainly a signal that results in better grading or access to information. In the extended version that we favor, cultural mobility also offers the opportunity for students from lower social classes to adopt the values, knowledge, and skills of the dominant social classes, including academic achievement.

Nevertheless, it should also be kept in mind that our findings are affected by some methodological and conceptual limitations. First, some covariates included in the multilevel linear regression models (e.g., the number of books in the home or the student's reported reading behavior), were not measured on an interval scale level. However, for the ease of model specification and interpretation, we assumed a linear relation between these covariates and reading comprehension. Second, our data were affected by a substantial amount of missing data and sample attrition. Therefore, our results could be biased if the data were not missing at random. Finally, we should be careful about assigning causal status to the reported effects. As only observational data were used, we are unable to exclude the existence of further unobserved or disregarded variables that might explain the relations we found between parents' cultural capital and students' reading achievement. Therefore, further research is needed to explain the mechanisms of social reproduction and mobility as well as differences in these mechanisms between students of different ethnic-cultural backgrounds.

References

Aschaffenburg, K., & Maas, I. (1997). Cultural and educational careers. The dynamics of social reproduction. *American Sociological Review, 62*, 573-587. Retrieved from http://www.asanet.org/journals/asr/index.cfm

Barone, C. (2006). Cultural capital, ambition and the explanation of inequalities in learning outcomes: A comparative analysis. *Sociology, 40*, 1039-1058. doi:10.1177/0038038506069843

Becker, B. (2010). The transfer of cultural knowledge in the early childhood: Social and ethnic disparities and the mediating role of familial activities. *European Sociological Review, 26*, 17-29. doi:10.1093/esr/jcn081

Becker, M., Lüdtke, O. Trautwein, U. Köller, O., & Baumert, J. (2012): The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology, 104*, 682-699. doi:10.1037/a0027608

Bouchard, T. J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212*, 1055-1059. doi:10.1126/science.7195071

Bourdieu, P. (1974). Cultural reproduction and social reproduction. In R. Brown (Ed.), *Knowledge, education and cultural change* (pp. 71-112). Tavistock Publications.

Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgement of Taste*. London: Routlegde.

Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241-258). New York: Greewood Press.

Bourdieu, P., & Passeron, J.-C. (1977). *Reproduction in Education, Society and Culture*. London: Sage.

Cheung, S. Y., & Andersen R. (2003). Time to read: Family resources and educational outcomes in Britain. *Journal of Comparative Family Studies, 34*, 413-433. Retrieved from http://soci.ucalgary.ca/jcfs/issues

Chiswick, B. R. (1978). The effect of Americanization on the earnings of foreign-born men. *Journal of Political Economy, 86*, 897-922. Retrieved from http://www.jstor.org/action/showPublication?journalCode=jpoliecon

Chiswick, B. R., & DebBurman, N. (2004). Educational attainment: analysis by immigrant generation. *Economics of Education Review, 23*, 361-379. doi:10.1016/j.econedurev.2003.09.002

Crook, C. J. (1997). *Cultural Practices and Socioeconomic Attainment: The Australian Experience*. Westport: Greenwood Press.

Cummins, J. (2003). Bilingual education. In J. Bourne & E. Reid (Eds.), *World yearbook of education: Language education* (pp. 3-19). London: Routledge.

De Graaf, N. D., De Graaf, P. M., & Kraaykamp, G. (2000). Parental cultural capital and educational attainment in the Netherlands: A refinement of the cultural capital perspective. *Sociology of Education, 73*, 92-111. Retrieved from http://soe.sagepub.com/content/by/year

DiMaggio, P. (1982). Cultural capital and school success: The impact of status culture participation on the grades of U.S. high school students. *American Sociological Review, 47*, 189-201. Retrieved from http://www.asanet.org/journals/asr/index.cfm

DiMaggio, P., & Mohr, J. (1985). Cultural capital, educational attainment, and marital selection. *American Journal of Sociology, 90*, 1231-1261. Retrieval from sociology.uchicago.edu/ajs/about.shtml

Esser, H. (2006). *Sprache und Integration. Soziale Bedingungen und Folgen des Spracherwerbs bei Immigranten*. Frankfurt a. M, New York: Campus.

Friedberg, R. M. (2000). You can't take it with you? Immigrant assimilation and the portability of human capital. *Journal of Labor Economics, 18*, 221-251. doi:10.1086/209957

Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology, 28*, 1096-1105. doi:10.1037/0012-1649.28.6.1096

Helberger, C. (1988). Eine Überprüfung der Linearitätsannahme der Humankapitaltheorie. In H.-J. Bodenhüfer (Ed.), *Bildung, Beruf, Arbeitsmarkt* (pp. 151–170). Berlin: Duncker & Humblot.

Hox, J. J. (2002). *Multilevel analysis. Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Kristen, C. (2008). Schulische Leistungen von Kindern aus Türkischen Familien am Ende der Grundschulzeit. Befunde aus der IGLU-Studie. In F. Kalter (Ed.), *Immigration und Integration. 48. Sonderheft der Kölner Zeitschrift für Soziologie und Sozialpsychologie* (pp. 230-251). Wiesbaden: VS-Verlag für Sozialwissenschaften.

Kristen, C., & Granato, N. (2007). The educational attainment of the second generation in Germany. Social origins and ethnic inequality. *Ethnicities, 7,* 343-366. doi:10.1177/1468796807080233

Lareau, A., & Weininger, E. B. (2003). Cultural capital in educational research. A critical assesment. *Theory and Society, 32,* 567-606. Retrieved from http://www.springerlink.com/content/0304-2421/

Lenhard, W., & Schneider, W. (2005). *ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässler.* Göttingen: Hogrefe.

Leopold, L., & Shavit, Y. (2013). Cultural capital does not travel well: Immigrants, natvies and achievement in Israeli schools. *European Sociological Review, 29,* 450-463. doi:10.1093/esr/jcr086

Lorenz, C. (2011). *Diagnostische Kompetenz von Grundschullehrkräften. Strukturelle Aspekte und Bedingungen.* Bamberg: University of Bamberg Press.

McElvany, N., Becker, M., & Lüdtke, O. (2009). Die Bedeutung familiärer Merkmale für Lesekompetenz, Wortschatz, Lesemotivation und Leseverhalten. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 41,* 121-131. doi:10.1026/0049-8637.41.3.121

Müller, A. G., & Stanat, P. (2006). Schulischer Erfolg von Schülerinnen und Schülern mit Migrationshintergrund: Analysen zur Situation von Zuwanderern aus der ehemaligen Sowjetunion und aus der Türkei. In J. Baumert, P. Stanat & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen. Vertiefende Analysen im Rahmen von PISA 2000* (pp. 221-255). Wiesbaden: VS Verlag für Sozialwissenschaften.

Nauck, B., Diefenbach, H., & Petri, K. (1998). Intergenerationale Transmission von kulturellem Kapital unter Immigrationsbedingungen. Zum Bildungserfolg von Kindern und Jugendlichen aus Immigrantenfamilien in Deutschland. *Zeitschrift für Pädagogik, 44,* 701-722. Retrieved from http://www.beltz.de/de/paedagogik/zeitschriften/zeitschrift-fuer-paedagogik/archiv.html

Nezlek, J. B., Schröder-Abé, M., & Schütz, A. (2006). Mehrebenenanalysen in der psychologischen Forschung. *Psychologische Rundschau, 57,* 213-223. doi:10.1026/0033-3042.57.4.213

Petrill, S. A., Deater-Deckard, K., Schatschneider, C., & Davis, C. (2005). Measured environmental influences on early reading: evidence from an adoption study. *Scientific Studies of Reading, 9,* 237-259. doi:10.1207/s1532799xssr0903_4

Plomin, R., & Spinath, F. M. (2004). Intelligence: Genetics, genes, and genomics. *Journal of Personality and Social Psychology, 86,* 112-129. doi:10.1037/0022-3514.86.1.112

Rodríguez-Brown, F. V. (2011). Family literacy. A current view of research on parents and children learning together. In M. L. Kamil, P. D. Pearson, E. Birr Moje & P. P. Afflerbach (Eds.), *Handbook of reading research* (pp. 726-754). New York, NY: Tayler & Francis.

Stanat, P. & Christensen, G. (2006). *Where immigrant students succeed – a comparative review of performance and engagement in PISA 2003.* Paris: OECD.

Stanat, P., & Edele, A. (2011). Migration und soziale Ungleichheit. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Eds.), *Empirische Bildungsforschung. Gegenstandsbereiche* (pp. 181-192). Wiesbaden: VS Verlag für Sozialwissenschaften.

Strand, S. (1999). Ethnic group, sex and economic disadvantage: associations with pupils' educational progress from baseline to the end of key stage 1. *British Educational Research Journal, 25*, 179-202. doi:10.1080/0141192990250204

Sullivan, A. (2001). Cultural capital and educational attainment. *Sociology, 35*, 893-912. doi: 10.1177/0038038501035004006

von Maurice, J., Artelt, C., Blossfeld, H.-P., Faust, G., Rossbach, H.-G., & Weinert, S.(2007). *Bildungsprozesse, Kompetenzentwicklung und Formation von Selektionsentscheidungen im Vor- und Grundschulalter: Überblick über die Erhebungen in den Längsschnitten BiKS-3-8 und BiKS-8-12 in den ersten beiden Projektjahren*. Retrieved from PsyDok: http://psydok.sulb.uni-saarland.de/volltexte/2007/1008

Weber, M. (1922/1978). *Economy and Society: An Outline of Interpretive Sociology*. (G. Roth & C. Wittich, Trans.). Berkeley. CA: University of California Press.

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 - Revision - (CFT 20-R) mit Wortschatztest und Zahlenfolgentest - Revision (WS/ZF-R)*. Göttingen: Hogrefe.

Wu, M. L., Adams, R. J., Wilson, M., & Haldane, S. A. (2007). *ACER ConQuest version 2.0: generalised item response modelling software*. Camberwell: ACER Press.

Chapter 7

7 Interest in Language Arts and Reading Competence in Secondary School

Irene M. Schurtz, Tobias Dörfler, Maximilian Pfost, and Cordula Artelt

Summary

Over the last 30 years, students' interests have increasingly been taken into account to explain individual differences in reading competence. In particular, the impact of students' interest in reading during preschool and primary school has been a topic of research. Fewer studies exist for students in secondary school, and only a limited number of studies have taken into account the development of the interests of secondary school students or have analyzed the impact of object-related individual interests on reading competence. In the present chapter, we address the missing link in this area of research by analyzing how students' interest in language arts and students' reading competence are related to each other in the first 2 years of secondary school. We found no direct effect of students' interest in

Author Note

Irene M. Schurtz,
Department of Educational Research, University of Bamberg, Germany;

Tobias Dörfler,
Institute of Psychology, University of Education, Heidelberg, Germany;

Maximilian Pfost,
Department of Educational Research, University of Bamberg, Germany;

Cordula Artelt,
Department of Educational Research, University of Bamberg, Germany.

This research was supported by Grant WE 1478/4-1 and AR 301/9-1 from the German Research Foundation (DFG)

Correspondence concerning this chapter should be addressed to Irene M. Schurtz, Department of Educational Research, University of Bamberg, Markusplatz 3, 96045, Bamberg, Germany. E-Mail: irene.schurtz@uni-bamberg.de

language arts on their reading competence, but we did find an indirect influence that was mediated by the amount of time that students spent reading. In-depth analyses comparing the time spent reading across different types of texts show that this indirect influence can be traced back to the amount of time spent reading narrative texts. Moreover, these results do not differ by gender, immigration background, or type of school. Finally, our analyses emphasize that the development of a student's interest in language arts and the student's reading competence are bidirectionally related to each other.

Research on the development of motivation and achievement has shown that relations between these two constructs are best described as complex and multifaceted. Thereby, over the last 30 years, interests have increasingly been taken into account when formulating explanations for the development of students' reading competence (cf. Hidi, Renninger, & Krapp, 2004; Wigfield & Asher, 1984). The present chapter focuses on the concept of interest in language arts as an object-related individual interest and analyzes its relation to the development of students' reading competence.

Reading Competence and Interest in Language Arts – Theoretical Conceptions and Developmental Perspectives

Being able to read represents a core competence in everyday life as dealing efficiently with written text is fundamental for citizens living in modern societies around the world (OECD, 2003). Reading competence, in particular, refers to the ability to formulate a coherent representation of a text. The act of reading itself is a complex one, which covers subprocesses across the different levels of words, sentences, and text. In order to create a coherent representation of a text, the reader needs to apply - more or less consciously - general world knowledge, syntactic knowledge, specific content knowledge, and metacognitive knowledge (Graesser, Millis, & Zwaan, 1997; Kintsch, 1998). Thus, reading competence represents the result of an interactive process between the reader and the text (Artelt, et al., 2005; Kintsch, 1998).

The development of students' literacy generally begins within the context of the family, and some students are already able to read and write when they enter primary school (Baker & Scher, 2002; Hurrelmann, 2004). However, for the majority of students, learning to read and write in a systematic manner begins in primary school. Whereas basic reading skills primarily develop in preschool and primary school, the ability and routine to draw inferences, create meaning from larger units of text, as well as the competent use of text develops mainly in Grade 4 and above (cf. Chall, 1983; McElvany & Becker, 2010). In order to become more and more familiar with the act of reading and the demands of text comprehension, to improve one's reading skills, and to develop a repertoire of (meta)cognitive reading strategies, students must keep encountering written text and must spend a lot of time reading (cf. Paratore, Cassano, & Schickedanz, 2011; Pfost, Dörfler, & Artelt, 2010). Thus, in addition to the reader's cognitive skills and prior knowledge, the role of a reader's level of motivation has received increased attention. Beginning with the work of Paris, Lipson, and Wixson (1983), the reader's skill and will to read began to be regarded as complementary. Researchers have thus increasingly been taking the reader's interest into account when explaining literacy development (cf. Miller & Faircloth, 2009).

Students' Interest in Language Arts

Referring to the Person-Object Conception of Interest (Krapp, 2002), interest is regarded as a relational construct that represents a particular relationship between a person and an object. This object- or content-specificity is the main factor that distinguishes interest from other motivational concepts (e.g., intrinsic motivation; Hidi & Ainley, 2002). Accordingly, Krapp (2002) points to three general structural components that describe a particular interest: first, the concrete topic of interest, which represents a certain domain of knowledge; second, specific activities that are connected to the object of interest and in which individuals are engaged when working on interest-related tasks; finally, real objects toward which the specific interest is directed. According to these three components, we may characterize interest in language arts in the following way: the German language and German literature are regarded as the topics or domains of interest. For students who are interested in language arts, reading can be regarded as one of the specific activities, and books can be seen as the typical objects of interest. Furthermore, interests are characterized by feeling- and value-related aspects, meaning that interest-related actions and contents have a subjective significance for the person and that the person likes to encounter them. Due to the positive feelings and significant personal value connected to this object or content, interested persons also generally have a tendency to enlarge their knowledge about the topic of interest and thus to improve their corresponding competencies (Krapp, 2000, 2002; Schiefele, 1999). These theoretical considerations are in line with research findings that have indicated that the connection between interest and achievement seems to get stronger as students grow older (Denissen, Zarrett, & Eccles, 2007; Schiefele, Krapp, & Winteler, 1992; Wigfield, et al., 1997). Finally, a person's interests can be divided into situational and individual components. Whereas situational interest describes a current engagement that occurs in and is created by a particular situation, individual interest depicts the dispositional structure of a person with related effects that tend to be long-lasting (Hidi & Renninger, 2006; Krapp, 2002). Correspondingly, before developing a dispositional interest, a person has to experience situational interest in a particular situation. Only if the engagement in a particular situation persists will the person be likely to develop an interest as a dispositional structure. Thus, to develop a long-lasting and profound individual interest, it is necessary to have the opportunity and will to re-engage in the interest-related activities (Hidi, et al., 2004; Hidi & Renninger, 2006; Renninger, 2000).

Following these theoretical considerations, the relation between interests and a person's competencies can be described in two different ways. On the one hand, competencies can be regarded as preconditions for the development of an interest. Due to the fact that a person's feeling of competence leads to positive feelings, the person is likely to develop an interest in topics that are related to activities in which he or she feels competent (cf. Daniels, 2008; Deci & Ryan, 2000; Krapp, 2005). On the other hand, it is assumed that people who have developed an interest in a particular domain tend to improve their interest-related competencies. Accordingly, competencies that develop through performing interest-related actions can also be regarded as consequences of a person's interest (cf. Krapp, 2000, 2002; Schiefele, 1999). Accordingly, because reading represents an interest-related activity, students' reading competence can be regarded as a predictor as well as an outcome of students' interest in language arts.

Students' Interest in Language Arts and Reading Competence – Previous Research Findings

For students in preschool and primary school, research findings have mainly indicated a positive effect of students' reading interest on their literacy development (e.g., Kirby, Ball, Geier, Parilla, & Wade-Wooley, 2011; Torppa, et al., 2007). With regard to secondary school students, comparable results have been reported concerning the positive relation between students' reading interest and their literacy (Möller & Schiefele, 2004). Furthermore, there is evidence that students' reading interest positively affects the amount of extracurricular reading that students do and thus their engagement in the reading process (e.g., Guthrie, Wigfield, Metsala, & Cox, 1999; McElvany, Kortenbruck, & Becker, 2008). As has been shown in various studies, the amount of reading, in turn, has a positive impact on reading competence (e.g., Anderson, Wilson, & Fielding, 1988; Guthrie, Schafer, & Huang, 2001; Pfost, et al., 2010). Thus, the positive influence of students' reading interest on their reading competence can be explained in part by increases in the reading practice of students who are interested in reading (Guthrie, et al., 1999; McElvany, et al., 2008; Wigfield & Guthrie, 1997). However, many of the studies that have analyzed students' interest in reading have shown some kind of design-based limitations such as using only teachers' (e.g., McKenna, et al., 1995) or parents' (e.g., Torppa, et al., 2007) reports to measure students' reading interest or by relying solely on cross-sectional data (e.g., Möller & Schiefele, 2004). The latter generally leads to an overestimate of the relation between students' interests and reading competence. Furthermore, the direction of influence remains unclear. Despite these limitations, from both a theoretical and an empirical point of view, there is reason to assume that in order to become a good reader, it is necessary to have reading-related skills at one's disposal but also to be willing to read. Previous research on students' interests and reading competence has revealed that empirical findings need to be distinguished according to the particular conceptualizations of interest they use. Especially when examining preschool and primary school students, previous studies have focused primarily on students' interest in reading. Although interest in language arts and interest in reading are both regarded as domain-specific interests, they represent different motivational concepts with regard to the Person-Object Conception of Interest (Krapp, 2002). According to Rheinberg (1998), interest in reading represents an activity-related motivation because the impulse to engage in this certain activity lies in the activity itself. By contrast, interest in language arts is an object-related motivation due to the fact that the impulse to perform a certain activity lies in a particular object that is related to this activity. There is some discussion in the literature indicating that only object-related motivation should be regarded as an interest due to the fact that only this type of motivation fulfills the theoretical assumptions needed to distinguish between a person, an object of interest, and the interest-related action that connect them (e.g., Krapp, 2002; Rheinberg, 1998; Schiefele & Schiefele, 1997). Given that students' interest in language arts is theoretically distinct and separable from their interest in reading, the empirical findings that have been reported thus far concerning the relation between interest in reading and reading achievement are not directly generalizable to students' interest in language arts. Previous empirical findings concerning students' interests and the impact of these interests on students' reading competence in secondary school have mainly focused on their topic interests which covers the triggered interest when a particular topic is presented (e.g., Ainley, Hidi, & Berndorff, 2002; Hidi, 2001). As an example, Renninger (1992) reported that fifth- and sixth-grade students who were interested in a certain topic of a text read this text more accurately and were able to recall more information from it than students who were not interested in this topic. Moreover, Schaffner, Schiefele, and Schneider (2004) found a significant positive relation between topic interest and reading comprehension for 15-year-old students in Germany. Nevertheless, given that topic interest was measured as an interest in the particular topic of the texts that were used to measure students' reading competence, these results may have been influenced by individual or situational interest (Ainley, et al., 2002; Hidi, 2000, 2001). Accordingly, these empirical findings are also not directly generalizable to the influence of students' individual interest in language arts because they refer to a different theoretical conception of interest. Thus, whether students' interest in language arts impacts students' reading competence in the beginning of secondary school remains an open question.

There is also evidence for the opposite effect: Reading competence might not be just an outcome of students' interest; it might also predict it. Corresponding research findings have indicated that in order to feel competent, students need to receive individual feedback on their skills and successes as well as to experience an optimal fit between their individual competencies and the requirements of the task. This feeling of competence, in turn, leads to positive feelings and promotes the development of students' interests. Thus, students who feel competent as readers are expected to enjoy the act of reading and thus be more likely to develop an interest in reading-related domains of interest (cf. Becker, McElvany, & Kortenbruck, 2010; Deci & Ryan, 2000; Daniels, 2008). However, there seem to be no studies that have used a longitudinal design to analyze reciprocal effects between students' reading competence and interest in secondary school (cf. Denissen, et al., 2007; Retelsdorf, Köller, & Möller, 2011). In the domain of mathematics, however, Marsh, Trautwein, Lüdtke, Köller, and Baumert (2005) found a bidirectional link between interests and achievement for students in Grade 7. Furthermore, their results indicate a smaller influence of achievement on interests than the opposite path, thus suggesting that a strong performance in a certain domain is not sufficient for developing an interest in this domain (cf. Renninger, Ewan, & Lasher, 2002). Taken together, students' reading competence can be seen as a necessary but not sufficient condition for developing an interest in language arts.

Finally, when analyzing the development of students' interest across the secondary school years, the finding that interests during this time period tend to decrease has to be taken into account (e.g., Daniels, 2008; McElvany, et al., 2008; Lüftenegger, et al., 2012; Wigfield, et al., 1997). This decline is often interpreted as a process of interest differentiation that begins in secondary school. Thus, whereas young children show a universal interest in nearly all activities, older students begin to develop domain-specific interests. This effect results in the persistence of high levels of interest in some specific domains, whereas for the same students, decreasing interest levels can be found in other domains. As a consequence of such a process of differentiation, decreases in interest scores on average are to be expected and have been observed several times (e.g., Daniels, 2008; Denissen, et al., 2007; Wigfield, et al., 1997). Furthermore, these assumptions are in line with the aforementioned empirical finding that the relation between (reading) interest and achievement seems to grow stronger as students grow older (Denissen, et al., 2007; Schiefele, et al., 1992; Wigfield, et al., 1997). Thus, students who develop a domain-specific interest across the school years persist in engaging in the interest-related activities of this particular interest, thus improving the competencies that are related to this interest. The associated positive feelings and feedback lead in turn to a continuously growing interest, and thus this interest tends to grow (Krapp, 2000; Schiefele, et al., 1992).

Students' Interest in Language Arts and Reading Competence – Potential Moderating Variables

It seems worthwhile to ask whether structural differences occur across certain subpopulations of students with regard to the relation between interest and reading competence. For example, Denissen, Zarrett, and Eccles (2007) found a weaker relation between interest and achievement for girls than boys. The authors interpreted this finding to reflect the idea that boys are mainly socialized to do well in particular domains, whereas girls are socialized to do well across domains (see also Logan & Johnston, 2009). Accordingly, boys primarily participate in domains they enjoy, whereas girls participate in all domains regardless of their interests (Schiefele, et al., 1992). In addition, students tend to view the act of reading as a typically female one (Eggert & Grabe, 2003; Millard, 1997; Philip, 2008). During adolescence, students' interests tend to develop in accordance with gender stereotypes, thus leading to a pattern of girls being more interested in reading-related activities (Hidi, et al., 2004; Meece, Glienke, & Burg, 2006; Renninger, 2000). Finally, there is evidence for a higher initial level of reading motivation as well as reading competence for girls than for boys, a difference that can mainly be explained by the greater amount of reading practiced by girls (e.g., Artelt, Naumann, & Schneider, 2010; Baker & Wigfield, 1999; Mullis, Martin, Foy, & Drucker, 2012; Wigfield & Guthrie, 1997).

Another potential moderator that should be taken into account is the type of school that students attend. Because the separation of students into the different types of schools is mainly based on students' school performance, students with severe reading deficits are more likely to attend schools in the lower (Hauptschule) or middle academic tracks (Realschule). Moreover, due to different institutional learning environments, further increases in these competence differences are expected (cf.

Pfost & Artelt, Chapter 8, this volume). With regard to the development of reading motivation, Retelsdorf and Möller (2008) reported that students attending upper academic track schools showed a higher initial level of reading motivation as well as a smaller decrease in reading motivation in comparison to students from lower and middle academic track schools. However, the impact of the type of school on the initial level as well as on the development of the students' interests remains unclear.

Finally students' immigration background should be taken into account as performance in written and spoken language depends on this variable (e.g., Baumert & Schümer, 2001; Chudaske, 2012; Naumann, Artelt, Schneider, & Stanat, 2010) although this effect is mainly attributable to differences in the often lower socio-economic backgrounds of the families (Marks, 2005). Even though parents born in a foreign country often show high educational aspirations, they frequently lag behind with regard to the opportunities to promote their children in terms of reading competence (e.g., Baumert & Schümer, 2001; Merkens & Nauck, 1993; Stanat, Rauch, & Segeritz, 2010). These findings hold for primary as well as secondary school students. Thus, whereas the enhanced performance of students without an immigration background with regard to their reading competence is evident, the impact of students' immigration background on their interests remains unclear.

In summary, students' reading competence and their interests are still subject to change across the secondary school years. However, previous research has mainly focused on students in primary school and their interest in reading, whereas the few studies that have analyzed students in secondary school have primarily analyzed topic interests and competencies in mathematics, and/or they did not account for reciprocal effects. Moreover, there are only a few studies that have used a longitudinal design. With regard to research on differential developments, students' reading competence has been studied intensively, whereas studies analyzing the effect of moderating factors on students' interests have mainly focused on students' gender. For this reason, the present chapter will focus on an object-related individual interest: the students' interest in language arts and its relation to the development of reading competence. To do so, we used longitudinal data measured during the first 2 years of secondary school. We also looked for the existence of structural differences according to students' gender, type of school, and immigration background.

Research Questions

In the present chapter, the following research questions were addressed:

1. Do the initial levels of students' interest in language arts, reading competence, and amount of extracurricular reading differ according to their gender, the type of school they attend, or their immigration background?

The first research question asks whether and to what extent students' gender, type of school, and immigration background lead to differences in the initial levels of their interest in language arts, reading competence, and amount of reading.

2. Does students' interest in language arts impact their reading competence? If so, is this effect mediated by the amount of reading that students do and does this effect vary across groups?

The second question asks whether and to what extent students' interest in language arts impacts on their reading competence. Based on previous empirical findings, a positive impact of the students' interest in language arts on their reading competence was expected. Moreover, in line with the findings and assumptions with regard to the behavioral effects of interest (Guthrie, et al., 1999; McElvany, et al., 2008; Krapp, 2000), we expected that this influence would be mediated by the amount of reading that students do: Interest in language arts should lead to large amounts of reading, which in turn should result in an increase in reading competence. Furthermore, we tested for structural differences in this relation by taking into account students' gender, immigration background, and the type of school as potential moderating factors.

3. Is there a connection between the development of students' reading competence and the development of the students' interest in language arts between Grade 5 and Grade 6?

As outlined above, in addition to being an outcome of a student's interest in language arts, reading competence can also be viewed as a predictor of the development of this interest. As the development of this interest is regarded as being strongly connected to feelings of competence, students with a below-average development of reading competence should experience more negative feelings while reading, leading to a decreasing interest in language arts. Thus, rather than focusing on a unidirectional model of influence, reciprocal effects were considered.

Method

Design and Participants

All analyses were based on data from $N = 1{,}631$ students who participated in the BiKS-8-14 panel study in Grades 5, 6, and 7 (assessment waves 4, 5, and 6 of the study; cf. Lorenz, Schmitt, Lehrl, Mudiappa, & Roßbach, Chapter 2, this volume). These 1,631 students (865 girls, 766 boys) attended 62 different secondary schools with an average of 23 students participating per school. In total, 979 (60.0%) of these students attended upper academic track schools, 308 (18.9%) middle academic track schools, and 344 (21.1%) lower academic track schools. The average age of the students in Grade 5 was 11.2 years ($SD = 0.5$). With regard to immigration background, the sample contained 226 (15.7%) students with one or two parents born abroad.

Measures

Interest in language arts. Interest in language arts was measured by a student questionnaire in Grades 5 and 6. The emotional and value-related aspects of the construct of interest were assessed by two items ("Reading and writing German texts by myself is great fun for me"; "It is important to me to become familiar with the German language and literature"). A third item measured whether and the extent to which students were willing to engage in interest-related activities during their spare time ("I am willing to use some of my spare time to get to know the German language and literature better"). The items were adapted from the BIJU study (Baumert, Gruehn, Heyn, Köller, & Schnabel, 1997) and were answered on a 5-point scale: 1 = *not at all*, 2 = *a little*, 3 = *moderately*, 4 = *fairly*, and 5 = *very much*. At both waves of assessment, the reliability of the scale was acceptable, especially when considering the small number of items used (Grade 5: Cronbach's $\alpha = .66$, Grade 6: Cronbach's $\alpha = .76$).

Reading competence. In Grade 5, reading competence was assessed by a sample of six short texts with a total of 43 multiple-choice items developed by the BiKS research group (Karing, et al., in prep.). Students had to read a given text, search relevant information, and generate more or less demanding inferences from the text to answer the given multiple-choice items. In Grade 6, three texts with a total of 31 multiple-choice items were used. Finally, in Grade 7, again, three texts with 26 multiple-choice items were used. For the three points of measurement, a common item design with nonequivalent groups/anchor-item test design was applied (Holland, Dorans, & Peterson, 2007; Kolen & Brennan, 2004); this allowed the estimation of students' reading competence to be placed on a common metric within an IRT framework. Item difficulty parameters for the same items across different assessments were set to be equal. In a first run, for the items on the Grade 5 reading competence test, the item difficulty parameters were estimated with a 1-parameter Rasch model by using the ConQuest software package (Wu, Adams, Wilson, & Haldane, 2007). The model was identified by setting the mean of the item difficulty parameters to zero. Item difficulty parameters of the Grade 6 and Grade 7 reading competence tests were estimated in subsequent second/third runs using the fixed item difficulty parameters from the foregoing point of measurement. Individual students' abilities were estimated by weighted likelihood estimates (WLEs) for every point of measurement. The reliabilities (WLE reliability) of the reading competence measures for all assessments were satisfactory (Grade 5 reliability = .78, Grade 6 reliability = .77, Grade 7 reliability = .76).

Time spent reading. The time spent in extracurricular reading was measured in telephone interviews with the students' parents in Grades 5, 6, and 7. Using an open scale, parents were asked to indicate how many hours per week their child reads for fun. Outliers were adjusted to a maximum of 20 hours per week, which approximately equals three standard deviations above the mean (cf. Pfost, et al., 2010).

Extracurricular reading behavior. Finally in Grade 7, extracurricular reading behavior was assessed by directly asking the students. Students were asked to indicate on a four point scale (1 = *almost never or never*, 2 = *several times a month*, 3 = *several times a week*, and 4 = *several times a day*) how often they read outside school. The ratings concerning the question ("How often do you read outside school...?") were asked separately for different types of text. The subsequent text types were used in this chapter: journals or newspapers; comics; novels, stories, or tales; and nonfiction books (e.g., technical or science).

Statistical Analyses

All statistical analyses were computed using SPSS 19 and Mplus 6.11 (Muthén & Muthén, 1998-2010). The first research question was examined by computing ANOVAs as well as standardized effect sizes using SPSS. The second research question was analyzed by applying structural equation modeling using the Mplus command *type = complex* to take the nested data structure into account. The full information maximum likelihood (FIML) estimation option in Mplus was used to handle missing data (cf. Preacher, Wichman, MacCallum, & Briggs, 2008). The percentage of missing data on the variables used in the following analyses varied between 0.1% (interest in language arts and reading competence in Grade 5) and 29.6% (reading competence in Grade 7).

To evaluate whether all the path analyses had to be computed as multigroup comparisons or whether it was sufficient to take the potential moderating variables into account as covariates, we tested for the existence of structural differences depending on the type of school, students' gender, and their immigration background. To do so, multigroup comparisons using the Satorra-Bentler-scaled chi-square difference test (Bryant & Satorra, 2011) were conducted to compare the adequacy of different equality constraints. In the first most restrictive model, the intercepts, variances, covariances, and regression paths were set equal between the comparison groups, thus suggesting that the particular grouping variable had no differential impact on the model variables. In the second model, the equality constraint of the intercepts was removed from the model, thus assuming that the intercepts varied between the comparison groups. In the third model of multigroup comparisons, the variances were additionally freely estimated, thus assuming that the model variables revealed group-specific variances. In the fourth model, the constraint of equal regression paths was additionally set free, thus allowing group differences in the structure of the relations of the model variables. In the fifth model, the covariances were also freely estimated. Thus, to test for the existence of structural differences depending on the potential moderating variables, the fourth model was of particular importance. A significant improvement in model fit from the third to the fourth model would reveal the existence of structural group differences. The model fit of all path analyses was evaluated by referring to three goodness-of-fit indices: The root mean square error of approximation (RMSEA), the chi-square test, and the comparative fit index (CFI; Preacher, et al., 2008). Models with RMSEA values of .05 or less, nonsignificant chi-square values, and CFI values above .95 were deemed acceptable (Hooper, Coughlan, & Mullen, 2008; Hu & Bentler, 1999).

The third research question was addressed by running a repeated-measures analysis of variance using SPSS. Moreover, these results were additionally tested by computing difference scores using Mplus.

Results

Descriptive Statistics and Correlations

The mean scores and standard deviations of all measures are presented in Table 1.

Table 1. Descriptive Statistics: Interest in Language Arts, Reading Competence, Time Spent Reading, and Reading Behavior

	M	*SD*	*Min*	*Max*	*N (Miss.)*
Grade 5					
Interest in language arts	3.18	0.87	1	5	1630 (1)
Reading competence (WLEs)	0.729	0.770	-	-	1629 (2)
Time spent reading	4.08	3.78	0	20	1429 (202)
Grade 6					
Interest in language arts	2.84	0.95	1	5	1409 (222)
Reading competence (WLEs)	1.076	0.957	-	-	1330 (301)
Time spent reading	5.00	4.46	0	20	1297 (334)
Grade 7					
Reading competence (WLEs)	1.275	1.126	-	-	1149 (482)
Time spent reading	4.74	4.06	0	20	1175 (456)
Reading behavior: Narrative texts	2.36	1.12	1	4	1271 (360)
Reading behavior: Nonfictional texts	1.55	0.80	1	4	1275 (356)
Reading behavior: Journals	2.42	0.92	1	4	1272 (359)
Reading behavior: Comics	1.67	0.94	1	4	1271 (360)

Note. Miss = Missing values; *Min* = theoretical minimum; *Max* = theoretical maximum; WLEs = weighted likelihood estimates.

With regard to the development of students' reading competence, the descriptive results indicated a steady increase from Grade 5 to Grade 7. The descriptive results of students' interest in language arts suggested a negative mean trend between Grade 5 and Grade 6. The average time spent reading increased from Grade 5 to Grade 6, whereas it decreased slightly in Grade 7. The cross-sectional descriptive analysis of students' reading behavior indicated that the students read journals and narrative texts more often than nonfictional texts and comics.

Research Question 1: Do the initial levels of students' interest in language arts, reading competence, and amount of extracurricular reading differ according to their gender, the type of school they attend, or their immigration background?

Additional descriptive analyses were computed with regard to differences on the potential moderating variables gender, type of school, and immigration background. To do so, the mean scores of students' interest and reading competence were compared between the different groups using ANOVAs (see Table 2).

Table 2. Average Scores and Standard Deviations of Students' Interest in Language Arts, Reading Competence, and Time Spent Reading in Grade 5 and Grade 6 Separated by Students' Gender, Immigration Background, and Type of School

	Gender			Type of school				Immigration background		
	Male	Female		Upper school track	Middle school track	Lower school track		No	Yes	
Grade 5	*M (SD)*	*M (SD)*	*p-value*	*M (SD)*	*M (SD)*	*M (SD)*	*p-value*	*M (SD)*	*M (SD)*	*p-value*
Interest in language arts	2.99 (0.90)	3.34 (0.81)	<.001	3.24 (0.82)	2.99 (0.92)	3.18 (0.94)	<.001	3.14 (0.89)	3.31 (0.82)	<.01
Reading compe-tence	0.70 (0.82)	0.76 (0.72)	n.s.	1.04 (0.66)	0.55 (0.60)	0.01 (0.65)	<.001	0.76 (0.76)	0.67 (0.78)	n.s.
Time spent reading	4.28 (3.74)	4.44 (3.77)	<.001	4.61 (3.89)	3.36 (3.67)	3.05 (3.14)	<.001	4.02 (3.65)	4.26 (3.95)	n.s.
Grade 6										
Interest in language arts	2.63 (0.95)	3.02 (0.91)	<.001	2.89 (0.95)	2.64 (0.97)	2.87 (0.89)	<.01	2.81 (0.95)	2.94 (0.95)	n.s.
Reading compe-tence	0.91 (0.98)	1.22 (0.91)	<.001	1.43 (0.87)	0.81 (0.72)	0.10 (0.69)	<.001	1.13 (0.95)	0.90 (0.98)	<.01
Time spent reading	4.28 (4.19)	5.65 (4.60)	<.001	5.73 (4.53)	4.07 (4.17)	3.46 (3.90)	<.001	5.03 (4.36)	4.53 (4.51)	n.s.

Note. The p-values indicate the significance of the mean score differences by students' gender/ type of school/ immigration background using ANOVAs. n.s. = statistically not significant.

Differential analyses for boys and girls indicated that girls reported a higher initial level of interest in language arts (Grade 5: $F(1, 1629) = 67.37$, $p < .001$, $d = 0.40$; Grade 6: $F(1, 1408) = 62.46$, $p < .001$, $d = 0.41$) and performed better on the reading comprehension test, although the effect was significant only in Grade 6 (Grade 5: $F(1, 1628) = 2.11$, $p = .15$, $d = 0.08$; Grade 6: $F(1, 1329) = 37.65$, $p < .001$, $d = 0.32$). Likewise, boys reported doing less reading outside school than girls (Grade 5: $F(1, 1428) = 15.12$, $p < .001$, $d = 0.20$; Grade 6: $F(1, 1296) = 31.20$, $p < .001$, $d = 0.31$). Differences in the initial levels of students' interest in language arts and reading competence according to the particular school track they attend were also analyzed by computing ANOVAs. To compare the three different academic school tracks with each other, planned contrasts were computed additionally. Results for the comparison of students' reading competence between the different school tracks indicated that students attending upper academic track schools performed better on the reading comprehension test than students attending middle (Grade 5: $t(1626) = 11.40$, $p < .001$, $d = 0.64$; Grade 6: $t(554) = 11.68$, $p < .001$, $d = 0.65$) and lower (Grade 5: $t(1626) = 25.30$, $p < .001$, $d = 1.34$; Grade 6: $t(445) = 24.24$, $p < .001$, $d = 1.39$) academic track schools. Students attending middle academic track schools, in turn, outperformed students from lower academic track schools (Grade 5: $t(1626) = 10.68$, $p < .001$, $d = 0.70$; Grade 6: $t(488) = 11.27$, $p < .001$, $d = 0.74$). With regard to interest in language arts, students attending upper academic track schools reported nearly the same level of interest as students attending lower academic track schools (Grade 5: $t(541) = 1.03$, $p = .31$, $d = 0.07$; Grade 6: $t(1406) = 0.21$, $p = .83$, $d = 0.02$). But both students attending upper (Grade 5: $t(472) = 4.25$, $p < .001$, $d = 0.29$; Grade 6: $t(1406) = 3.75$, $p < .001$, $d = 0.26$) and lower (Grade 5: $t(643) = 2.62$, $p = .009$, $d = 0.22$; Grade 6: $t(1406) = 2.86$, $p = .004$, $d = 0.24$) academic track schools reported a higher interest in language arts than students attending middle academic track schools. With regard to the amount of extracurricular reading, upper academic track school students read significantly more (more hours per week) during their spare time than students attending middle (Grade 5: $t(494) = 4.91$, $p < .001$, $d = 0.33$; Grade 6: $d = 0.37$, $t(450) = 5.43$, $p < .001$) and lower (Grade 5: $t(535) = 6.72$, $p < .001$, $d = 0.41$; Grade 6: $t(425) = 7.55$, $p < .001$, $d = 0.51$) academic track schools. However, the comparison between students attending middle and lower academic track schools showed no significant differences with regard to their amount of reading during spare time (Grade 5: $t(539) = 1.06$, $p = .288$, $d = 0.08$; Grade 6:

$t(482) = 1.66$, $p = .098$, $d = 0.14$). With regard to students' immigration background, we first analyzed whether students with one parent born abroad differed in their interest in language arts, reading competence, and time spent reading from students with two parents born abroad so that it would be necessary to differentiate between these two subgroups of students with immigration backgrounds. The analyses revealed significant differences only with regard to reading competence in Grade 5. Because there were almost no differences between these two immigration subgroups, the following analyses differed only between students with an immigration background (one or two parents born abroad) and without an immigration background (both parents born in Germany). Differential analyses between students with and without an immigration background revealed that students' immigration background made a difference inasmuch as students with an immigration background - compared to students whose parents were both born in Germany - achieved lower scores on the reading competence test although the difference was statistically significant only in Grade 6 (Grade 5: $F(1, 1441) = 2.31$, $p = .13$, $d = 0.12$; Grade 6: $F(1, 1189) = 8.47$, $p = .004$, $d = 0.24$). However, students with an immigration background reported a greater interest in language arts, although this difference was statistically significant only in Grade 5 (Grade 5: $F(1, 1442) = 6.79$, $p = .009$, $d = 0.20$; Grade 6: $F(1, 1260) = 2.99$, $p = .08$, $d = 0.14$). With regard to students' time spend reading during their spare time students with and without immigration background showed no differences in Grade 5 and 6 (Grade 5: $F(1, 1325) = 0.68$, $p = .41$, $d = 0.06$; Grade 6: $F(1, 1186) = 1.87$, $p = .17$, $d = 0.11$).

Taken together, these first differential analyses indicated substantial influences of the potential moderating variables gender and type of school on the outcome variables of interest, whereas with regard to students' immigration background, only minor effects were shown.

Table 3 depicts correlations between students' interest in language arts, reading competence, and time spent reading in Grades 5, 6, and 7, and indicates that all variables were positively correlated.

Table 3. Correlations between Interest in Language Arts, Reading Competence, and Time Spent Reading in Grades 5, 6, and 7

Measure	1	2	3	4	5	6	7	8
1. Interest in language arts (G5)	-							
2. Interest in language arts (G6)	.52*	-						
3. Reading competence (G5)	.07*	.07*	-					
4. Reading competence (G6)	.09*	.16*	.59*	-				
5. Reading competence (G7)	.12*	.13*	.55*	.64*	-			
6. Time spent reading (G5)	.19*	.20*	.29*	.28*	.32*	-		
7. Time spent reading (G6)	.17*	.19*	.24*	.28*	.30*	.52*	-	
8. Time spent reading (G7)	.18*	.24*	.21*	.28*	.31*	.52*	.55*	-

Note. $N = 1{,}631$; G = Grade.
$^{*}p < .05$.

A closer look at the depicted correlations reveals that interest in language arts was more highly correlated with time spent reading than with reading competence. Furthermore, reading competence showed higher correlations with students' amount of extracurricular reading than with their interest in language arts. Additional analyses were computed to take into account the particular text types that were read by the students (see Table 4).

Table 4. Correlations between Reading Competence, Time Spent Reading, and Reading Behavior in Grade 7

Measure	1	2	3	4	5	6
1. Reading competence	-					
2. Time spent reading	.33*	-				
3. Reading behavior: Narrative texts	.41*	.47*	-			
4. Reading behavior: Nonfictional texts	.06*	.09*	.18*	-		
5. Reading behavior: Journals	.09*	.07*	.16*	.17*	-	
6. Reading behavior: Comics	.03	.12*	.06*	.22*	.16*	-

$^{*}p < .05$.

These additional results in Grade 7 revealed that students' interest in language arts and reading competence were most highly correlated with the amount of narrative reading that the students did. Students' interest in language arts was also significantly correlated with the amount of reading of nonfictional texts and journals. However, the amount of reading of both text types was not correlated at all (for nonfictional texts) or was only moderately correlated (for journals) with students' reading competence in Grade 7. Last, the amount of comic reading was not related to interest in language arts or reading competence. In conclusion, the correlations provide a first hint for a potential indirect relation between students' interest in language arts and their reading competence which is mediated by the time students spend reading. Moreover, especially students' reading behavior of narrative texts was positively related to students' interest in language arts and reading competence.

Longitudinal Data Analysis

Research Question 2: Does students' interest in language arts impact their reading competence? If so, is this effect mediated by the amount of reading that students do and does this effect vary across groups?

In the first step of this analysis, the relations between interest in language arts and reading competence in Grade 5 and Grade 6 were tested for structural differences depending on school type, students' gender, as well as immigration background. To do so, a cross-lagged panel model (see Figure 1) was computed using type of school, gender, and immigration background as separate grouping variables in a multigroup model to compare the adequacy of different equality constraints. In the most restricted model, mean scores, variances, co-variances, and regression paths of the interest in language arts and reading competence in Grade 5 and Grade 6 were set equal, thus suggesting that the particular grouping variable had no differential impact on the model variables. The fit parameters indicated that the estimated coefficients did not fit the empirical data for all three grouping variables (see Table 5). Thus, in the second step, mean scores of the model variables were freely estimated, thus assuming that level differences existed between groups. Confirming the results of the first research question, this step resulted in a significant improvement in the model fit compared to the previous restricted model, thus indicating that all grouping variables had a differential impact on the mean scores of the model variables. Thereupon, the variances were also estimated freely. Whereas, when immigration background was used as the grouping variable, these model modifications did not improve the model fit further, the fit indices of the other two comparison models improved significantly when the variances were also estimated freely (Model 3). However, setting the estimation of the regression paths free in the fourth model did not result in a significant improvement in the model fit, thus indicating that the particular structural influences between the model variables did not differ between the categories of the grouping variables. In the same manner, the mediation model presented in Figure 2 was analyzed for the existence of structural differences. Again, differences occurred with regard to the initial levels of the model variables. However, there were no structural differences across the groups. Thus, based on these results, it is not necessary to compute multigroup comparison models but to use gender, type of school, and immigration background as covariates in a single group model.

Table 5. Examining Structural Differences according to Students' Gender, Type of School, and Immigration Background with Regard to the Relation of Students' Interest in Language Arts and Reading Competence in Grade 5 and Grade 6 (see Figure 1)

	Model 1	Model 2	Model 3	Model 4
Gender differences				
χ^2 value (df)	120.843 (14)	30.451 (10)	8.465 (6)	5.014 (4)
RMSEA	.097	.050	.022	.018
CFI	.925	.986	.998	.999
AIC	14433.204	14327.855	14309.484	14310.223
TRd [a] (delta df)	120.843 (14)	82,819 (4)	21.16 (4)	3.741 (2)
p-value	< .05	< .05	< .05	> .05
Differences by type of school				
χ^2 value (df)	768.067 (28)	56.337 (20)	28.516 (12)	20.528 (8)
RMSEA	.220	.058	.050	.054
CFI	.000	.951	.978	.983
AIC	14433.204	13726.117	13705.010	13707.074
TRd[a] (delta df)	768.067 (28)	931.403 (8)	25.92 (8)	7.952 (4)
p-value	< .05	< .05	< .05	> .05
Differences by immigration background				
χ^2 value (df)	32.371 (14)	13.487 (10)	12.440 (6)	7.405 (4)
RMSEA	.043	.022	.039	.034
CFI	.985	.997	.995	.997
AIC	12841.512	12825.062	12830.716	12830.082
TRd [a] (delta df)	32.371 (14)	16.814 (4)	1.828 (4)	5.303 (2)
p-value	< .05	< .05	> .05	> .05

Note. Model 1 = fixed mean values, variances, covariances, and regression paths; Model 2 = fixed variances, covariances, and regression paths; Model 3 = fixed covariances and regression paths; Model 4 = fixed covariances.
[a]Satorra-Bentler-scaled χ^2 difference test.

In the first structural equation model, the relation between interest in language arts and reading competence in Grade 5 and Grade 6 was analyzed (Figure 1) by specifying a cross-lagged path model controlling for the impact of the type of school, gender, and immigration background as covariates. According to the results, both constructs could be characterized as stable across the two assessments. Furthermore, the cross-lagged paths indicated that interest in language arts did not affect reading competence at the subsequent assessment, and students' reading competence in Grade 5 did not impact students' interest in language arts in Grade 6.

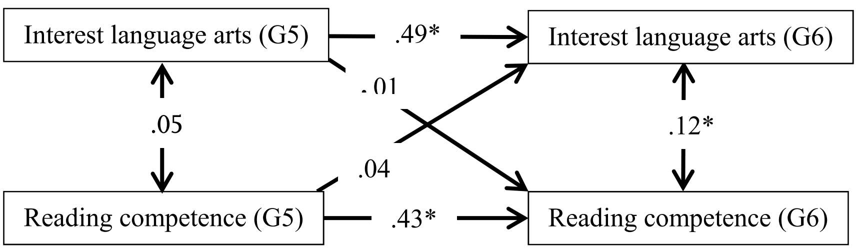

Figure 1. Associations between interest in language arts and reading competence in Grade 5 and Grade 6 taking into account students' gender, type of school, and immigration background as covariates (standardized path coefficients ß). $N = 1{,}631$; G = Grade. $^*p < .05$.

Putting these results together, we concluded that there was a significant relation between students' interest in language arts and reading competence controlling for background variables, but only in Grade 6. However, when controlling for Grade 5 reading competence, our analysis did not reveal a direct effect of students' interest in language arts in Grade 5 on their reading competence in Grade 6; thus, the research hypothesis related to the second question was not supported. Furthermore, the reciprocal effect of students' competencies on their interest in language arts was not significant.

Despite these findings, we analyzed whether there was at least a tendency toward an indirect effect of students' interests in language arts on their reading competencies mediated by students' extracurricular reading behavior. The corresponding path model is depicted in Figure 2. Controlling for students' reading competence in Grade 5, results for gender, type of school, and immigration background indicated a significant effect of interest in language arts on the amount of time students spent doing extracurricular reading. Again, students' interest in language arts in Grade 5 did not directly affect their reading competence in Grade 6. Extracurricular reading behavior, however, was positively related to the students' reading competence in Grade 6. Taken together, we found an indirect effect of interest in language arts in Grade 5 on students' reading competence in Grade 6 mediated by students extracurricular reading behavior ($\beta_{indirect} = .02$, $p < .001$).

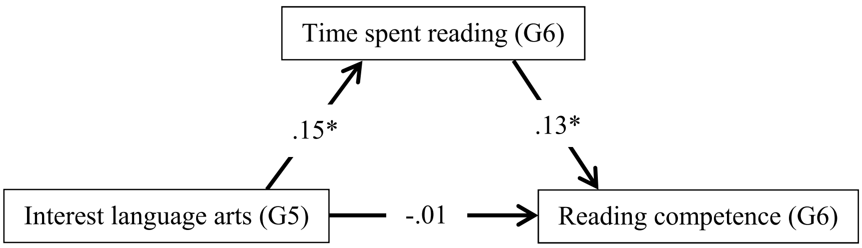

Figure 2. Indirect relation between interest in language arts in Grade 5, amount of reading in Grade 6, and reading competence in Grade 6 taking into account students' previous reading literacy in Grade 5, gender, type of school, and immigration background as covariates (standardized path coefficients ß). $N = 1{,}631$; G = Grade. $^{*}p < .05$.

Moreover, we were able to replicate and improve this connection by analyzing interest in language arts in Grade 6 and students' time spent reading as well as reading competence in Grade 7 accounting for the same background variables ($\beta_{indirect} = .03$, $p < .001$). Because we additionally measured how often students read different text types in Grade 7, we were able to analyze whether this indirect relation could be traced back to particular text types. The comparison between narrative texts, nonfictional texts, journals, and comics indicated that -according to the descriptive analyses - the indirect relation between interest in language arts and reading competence could be traced back to how often students read narrative texts. The specific indirect effect again increased to $\beta_{indirect} = .04$ ($p < .001$), controlling for prior performance (reading competence in Grade 6), gender, type of school, and immigration background. The other text types did not significantly connect these two constructs when controlling for the same background variables (nonfictional texts: $\beta_{indirect} = .004$, p = .521; journals: $\beta_{indirect} = .002$, p = .388; comics: $\beta_{indirect} = .001$, p = .521).

Research Question 3: Is there a connection between the development of students' reading competence and the development of the students' interest in language arts between Grade 5 and Grade 6?

The question of whether or not students' reading competence is related to their interest development was examined by using a repeated-measures analysis of variance. Therefore, the development of the students' interest in language arts from Grade 5 to Grade 6 was included as the dependent variable, whereas the development of their reading competence was included as the independent variable. To examine whether or not different courses of competence development are related to different courses of interest development, the students' development of reading competence was classified relative to all students in the sample: In a first step, students were classified into three groups, separately for Grade 5 and Grade 6, by defining students with a reading competence score of one standard deviation above the average as students with a *high relative reading competence* and students with a score of one standard deviation below the average as students with a *low relative reading competence*. Students in between these two boundaries were defined as students with an *average relative reading competence*. In a second step, students were classified as having an increased, a decreased, or a stable relative reading competence according to the change in their classification from Grade 5 to Grade 6. The frequency distribution of this new variable *grouped relative reading competence development* indicated that out of a total of 1,328 students who were included in the analyses, over half of the students (57.4%) had a stable relative reading competence from Grade 5 to Grade 6 (see Figure 3). Approximately one third of the students (33.4%) were classified as having an increasing relative reading competence, and 9.4% of the students had a decreasing relative reading competence.

Relative reading competence		Development of relative reading competence
Grade 5	Grade 6	Grade 5 to Grade 6
Medium	High	Increasing relative reading competence (33.4%)
Low	High	
Low	Medium	
High	High	Stable relative reading competence (57.4%)
Medium	Medium	
Low	Low	
High	Medium	Decreasing relative reading competence (9.4%)
High	Low	
Medium	Low	

Figure 3. Classification of the development of students' relative reading competence from Grade 5 to Grade 6. *N* = 1,328.

Results of the repeated-measures analysis of variance indicated a significant main effect of time, $F(1, 1324) = 157.10$, $p < .001$, part. $\eta^2 = .106$, indicating that on average, the developmental trend of the students' interest in language arts was decreasing. Whereas there was no main effect of grouped relative reading competence development on students' interest in language arts, $F(2, 1324) = 2.12$, $p = .12$, part. $\eta^2 = .003$, a significant interaction effect was found, $F(2, 1324) = 4.70$, $p = .01$, part. $\eta^2 = .007$. As shown in Figure 4, the decrease in interest in language arts was significantly accelerated for students in the decreasing relative competence group. Thus, a decreasing relative reading competence score was accompanied by a more markedly decreasing interest in language arts.

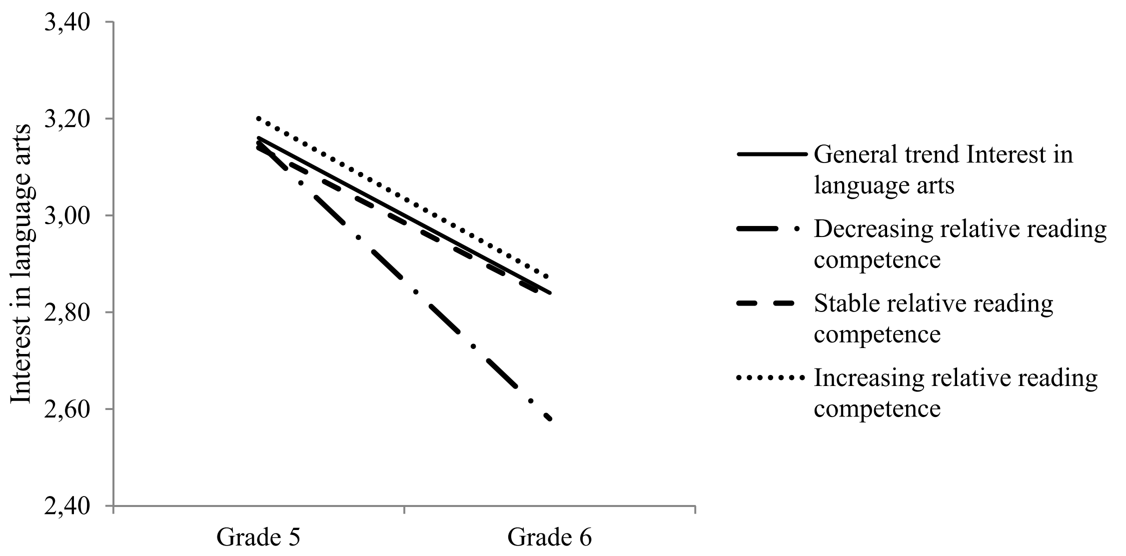

Figure 4. The development of students' interest in language arts depending on the students' development of relative reading competence classified into three groups. $N = 1{,}328$.

A change model was also computed to additionally test the second research question. The results reinforced the presented conclusions, indicating that the initial level of reading competence was not significantly correlated with interest in language arts, but that there was a small although significant correlation between the difference scores of the two constructs ($r = .09$, $p = .004$). Thus, the overall correlations also indicated that the particular developmental processes of the students' interest in language arts and reading competence from Grade 5 to Grade 6 influenced each other.

Discussion

The aim of this chapter was to analyze the relation between students' interest in language arts and the development of their reading competence in the first 2 years of secondary school. Contrary to our expectations, the presented results indicated no direct effect of students' interest in language arts on their reading competence. However, we were able to show an indirect connection between these two measures through the amount of extracurricular reading that the students did. Additional analyses showed that this indirect relation could be traced back to how often the students read narrative texts. Moreover, we analyzed whether or not these results held when controlling for students' gender, type of school, and immigration background. These analyses revealed unique differences in the initial levels of students' interest in language arts, reading competence, as well as time spent reading but no structural ones. With regard to the third research question, the results indicated that a relative decrease in reading competence was attended by a more pronounced decrease in the students' interest in language arts. Thus, the presented results supported a reciprocal relation between the development of students' reading competence and interest in language arts.

In contrast to previous studies that have examined the impact of interests on reading competence, we analyzed a domain-specific individual interest, which is more distal from the act of reading and the reading test than previously researched general interests in reading interest or topic interest. In the current study, interest in language arts was conceptualized as an object-related motivation and thus represented a complex construct of interest that contains the act of reading in only an instrumental way. This unique conception of interest could provide one possible explanation for why the direct link between students' reading competence and interest in language arts was not found in our study. Thus, students could be interested in language arts but might use other interest-related activities to engage in it (e.g., attending a lecture or playing language games). According to our results, students who are interested in language arts will show increases in their reading competence only if, due to this high interest in language arts, they also engage in large amounts of reading.

A second explanation refers to the particular situation in which measurements are taken. According to Köller, Baumert, and Schnabel (2000), students are mainly extrinsically motivated in the school context as well as when taking achievement tests. Thus, perhaps a student's particular interest would not additionally affect the student's achievement on the test because it is masked by the student's extrinsic motivation in the test situation, such as the motivation to achieve good marks or test scores. By contrast, in situations that are characterized by intrinsic motivation, students who do not have an interest in language arts would be expected to be less motivated to read, whereas interested students should read more continuously (Köller, et al., 2000). In addition, whereas students' spare time offers them the opportunity to engage in many different activities, school time is mainly characterized by a lack of choice with regard to activities. Thus, during spare time, students' interests can be operative in determining their actions and, as a result, may have a positive effect on students' competence development (Baker & Wigfield, 1999; Köller, et al., 2000; McElvany, et al., 2008). In accordance with this idea, we were able to show that students who were interested in language arts had significantly more often chosen to read in their spare time, thus improving their reading competence. Therefore, with regard to a long-lasting impact of students' interest in language arts, we expect that this interest will develop a direct effect on students' reading competence in the higher secondary school grades. This expectation is further underlined by the presented results, which revealed that correlations and influences became stronger across the two measurement points. These findings correspond to previous studies that found increasing correlations within interest domains and decreasing correlations between different ones as well as studies that found increasing relations between interests and reading competence over time which were interpreted to indicate a process of increasing consolidation of students' interest (Denissen, et al., 2007; Schiefele, et al., 1992; Wigfield, et al., 1997).

The roles of the proposed moderating variables (i.e., gender, type of school, and immigration background) were addressed by our first and second research questions. The analyses of these potential moderating variables indicated no structural differences beyond differences in the mean scores. Thus, according to the type of school, the reported results did not support the existence of institutional effects on the relation between students' interest in language arts and their reading competence. However, students attending upper academic track schools showed a greater interest in language arts and a higher reading competence, thus supporting Retelsdorf and Möller's (2008) findings. Analyses with regard to the influence of students' gender on the initial competence and interest levels supported the results that were noted from previous studies (e.g., Baker & Wigfield, 1999; Wigfield & Guthrie, 1997): Girls reported a greater interest in language arts, a greater amount of reading, and achieved a higher reading competence than boys did. However, our results emphasize that there are no structural differences according to students' gender. This finding contradicts prior research that indicated a weaker relation between achievement and interest for girls than for boys. These differential structural relations were interpreted to be the result of gender differences in the socialization process whereupon girls are socialized to do well in many different domains, regardless of their interests, whereas boys are socialized to do well in the particular domains in which they are interested (Denissen, et al., 2007; Schiefele, et al., 1992). With regard to the constructs that were analyzed in this chapter, it seems that the observed gender differences in reading competence can be traced back to gender differences in attitudes toward reading but not to the existence of gender-specific mechanisms that link students' interest in language arts to reading competence. The different attitudes may be caused by the previously mentioned gender stereotype (i.e., that the act of reading is mainly a female activity; Hidi, et al., 2004; Millard, 1997; Renninger, 2000). With regard to immigration background, the reported results support the findings of previous studies in indicating that students with an immigration background show a lower reading competence than students without an immigration background (Baumert & Schümer, 2001; Chudaske, 2012; Schleicher, 2006). The reported amount of extracurricular reading showed no significant differences with regard to the initial level. However, students with an immigration background reported a greater interest in language arts than students who did not have an immigration background, thus contradicting the previously presented indirect relation between these three constructs. Thus, a higher interest in language arts did not lead to more extracurricular reading as one of the interest-related actions. It is conceivable that students with an immigration background tend to focus more on the language-related aspects of this interest than on the literature-related aspects because language-related aspects are of more importance in their everyday lives. As a result, other interest-related actions (e.g., taking a language course, playing language games, or listening to CDs) could be of more importance to them when engaging in actions that are related to their interest in language arts. Another possible explanation concerns socially desirable responding. Students with an immigration background might think that they are expected to answer in a positive manner when questioned about culturally characterized behavior (Aschauer, 2009). The question of whether they are interested in the German language and literature could have triggered such positively biased behavior with regard to their answers. However, our findings are in line with previous ones, such that students with an immigration background tend to show a lower reading competence but also higher educational aspirations than students without an immigration background (Baumert & Schümer, 2001; Chudaske, 2012; Schleicher, 2006). Thus, apparently students with an immigration background have a high motivational potential but are not able to transfer this to concrete actions (Stanat, et al., 2010). Perhaps teachers have more difficulty in identifying and promoting the motivations and interests of students with immigration backgrounds in particular (Stanat, et al., 2010). But, further investigation is needed to explain these findings. To do so it would be interesting to consider not only the immigration background by country of origin but also immigration background by students' commonly used language at home. Due to the fact that reading competences are strongly connected to students' language abilities, further studies should take both measures into account.

Finally, with regard to the third research question, our results indicated that the development of students' reading competence and their interest in language arts are mutually dependent upon each other, though both constructs seemed only weakly related. Whereas there was a general trend toward decreasing interest in language arts from Grade 5 to Grade 6, students who successfully improved their reading competence showed a slightly smaller decrease in their interest in language arts. Furthermore, students with a decreasing relative reading competence showed a more pronounced negative trend in their interest in language arts. This significant interaction between the development of students' interest in language arts and their relative reading competence indicates that even though students with a high reading competence will not necessarily develop an interest in the German language and literature, students with relatively low competencies in interest-related activities are likely to turn away from those domains. Accordingly, reading competence seems to be a necessary but not sufficient precondition for students' interest in language arts. Thus, although we could not find significant cross-lagged paths between students' interest in language arts and reading competence, our results indicate a significant relation between the changes in the two constructs, consequently highlighting the importance of taking interest-related competencies into account when researching interest development (Daniels, 2008; Deci & Ryan, 2002; Ryan & Deci, 2000).

Limitations of the Study

The first limitation of our study concerns the measurement of the amount of time students spent doing extracurricular reading. This variable was based on information provided by the students' parents, who were asked to estimate how many hours per week their children spent doing extracurricular reading. One criticism of this measurement is that parents do not necessarily know how many hours their children spend reading per week, especially when the parents are working. Moreover, parents must differentiate between the number of hours their children read for fun and for school. Thus, the accuracy of these parental estimations remains unclear. Nevertheless, an additional analysis for which extracurricular reading behavior was self-reported by the students revealed comparable results.

Second, the presented results were not controlled for the influence of students' prior reading competence or their interest in language arts in Grade 4 because this information was not available for all students.

Third, we may criticize that only the first 2 years of secondary school were analyzed. Therefore, any additional development in the students' interest in language arts could not be pursued. Moreover, we still do not know whether or not the missing direct effect of students' interest in language arts on their reading competence would have been observed if the subsequent years of secondary school had been analyzed. Likewise, structural differences according to students' gender, immigration background, and the school track they attended might arise later when differences have become stronger and more consolidated.

Conclusion and Further Research Questions

In summary, we were able to show that although students' interest in the German language and literature did not directly affect their reading ability, there was an indirect relation that was mediated by the amount of extracurricular time the students spent reading, especially with regard to the reading of narrative texts. Moreover, this relation applied to students of both genders, students with and without immigration backgrounds, and students attending different types of schools. Finally, our results indicated a joint development of students' reading competence and interest in language arts and thus support the conclusion that these effects are reciprocal.

Even though the background variables that we considered did not influence the relation between students' interest in language arts and their reading competence, the impact of other moderating variables is still possible, and research on their effects should be pursued. For example, with regard to students' reading motivation, the reading behavior of their parents has been found to significantly influence the development of this motivational construct (e.g., Baumert & Schümer, 2001; Baker & Scher, 2002). Thus, the influence of the attitude of students' parents toward the German language and literature should be additionally examined as a potential moderating variable.

Furthermore, future research should examine the relation between students' interest in language arts and their reading competence over a longer period of time. Beginning in primary school, these analyses could provide important information about whether students' interest in language arts and reading competence develop in a reciprocal manner or whether one construct primarily influences the other. Although our results suggest a joint development of students' interest in language arts and reading competence, more extensive data are needed to replicate and broaden our findings. Nevertheless, the presented results demonstrate that it is worthwhile to analyze an object-related interest and its specific impact on students' reading competence.

References

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, 94, 545-561. doi: 10.1037//0022-0663.94.3.545

Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 23 (3)*, 285-303. Retrieved from http://www.jstor.org/stable/748043

Artelt, C., McElvany, N., Christmann, U., Richter, T., Groeben, N., Köster, J. (...) Ring, K., (2005). *Expertise: Förderung von Lesekompetenz*. Berlin: BMBF.

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: Waxmann.

Aschauer, W. (2009). Besonderheiten und Problemlagen der quantitativen Befragung bei MigrantInnen. In M. Weichbold, J. Bacher & C. Wolf (Eds.). *Umfrageforschung: Herausforderungen und Grenzen* (pp. 293-316). Wiesbaden: VS Verlag.

Baker, L., & Wigfield, A. (1999). Dimensions of children's motvitation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly, 34* (4), 452-477. Retrieved from http://www.jstor.org/stable/748216

Baker, L., & Scher, D. (2002). Beginning readers' motivation for reading in relation to parental beliefs and home reading experiences. *Reading Psychology, 23*, 239-269. doi: 10.1080/02702710290061346

Baumert, J., Gruehn, S., Heyn, S., Köller, O., & Schnabel, K.-U. (1997). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU) Dokumentation – Band 1, Skalen Längsschnitt I, Welle 1-4.* Berlin: Max-Plank-Institut für Bildungsforschung, Forschungsbereich Erziehungswissenschaften und Bildungssysteme.

Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In Deutsches PISA-Konsortium (Eds.). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 69-137). Opladen: Leske + Budrich.

Becker, M., McElvany, N., & Kortenbruck, M. (2010). Intrinsic and Extrinsic Reading Motivation as Predictors of Reading Literacy: A Longitudinal Study. *Journal of Educational Psychology, 102*, 773-785. doi: 10.1037/a0020084

Bryant, F. B., & Satorra, A. (2011). *Scaled difference chi-square testing*. Retrieved from http://www.statmodel.com/download/BryantSatorraInPressSEM2011.pdf.

Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill Book Company.

Chudaske, J. (2012). *Sprache, Migration und schulfachliche Leistung*. Wiesbaden: VS Verlag.

Daniels, Z. (2008). *Entwicklung schulischer Interessen im Jugendalter*. Münster: Waxmann.

Deci, E. L., & Ryan, R. M. (Eds.). (2002). *Handbook of self-determination research*. Rochester, NY: University of Rochester Press.

Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11* (4), 227-268.

Denissen, J. J. A., Zarrett, N. R., & Eccles, J. S. (2007). I like to do it, I’m able, and I know Iam: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development, 78* (2), 430-447.

Eggert, H., & Garbe, C. (2003). *Literarische Sozialisation* (2nd, Rev. ed.). Stuttgart: Metzler.

Guthrie, J. T., Schafer, W. D., & Huang, C. (2001). Benefits of opportunity to read and balanced instruction on the NAEP. *Journal of Educational Research, 94* (3), 145-162. doi: 10.1080/00220670109599912

Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading, 3* (3), 231-256. doi: 10.1207/s1532799xssr0303_3

Graesser, A.C., Millis, K.K., & Zwaan, R.A. (1997). Discourse comprehension. *Annual Reviews of Psychology*, 48, 163-189.

Hidi, S., & Ainley, M., (2002). Interest and adolescence. In F. Pajares & T. Urdan (eds.), *Academic motivation of adolescents* (pp. 247-276). Greenwich: Information Age Publishing.

Hidi, S. (2001). Interest, reading, and learning: Theoretical and practical considerations. *Educational Psychology Review, 13* (3), 191-209.

Hidi, S. (2000). An interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation. In C. Sonsone, & J. M. Harackiewicz (Eds.). *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 309-339). San Diego: Academic Press.

Hidi, S., Renninger, K. A., & Krapp, A. (2004). Interest, a motivational variable that combines affective and cognitive functioning. In D. Y. Dai & R. J. Sternberg (Eds.). *Motivation, Emotion, and cognition: Integrative perspectives on intellectual functioning and development.* London: Lawrence Erlbaum Associates.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41* (2), 111-127. doi: 10.1207/s15326985ep4102_4

Holland, P. W., Dorans, N. J., & Peterson, N. S. (2007). Equating test scores. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics 26. Psychometrics* (pp. 169-203). Amsterdam: Elsevier.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modeling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6* (1), 53-60. Retrieved from www.ejbrm.com

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. doi: 10.1080/10705519909540118

Hurrelmann, B. (2004). Sozialisation der Lesekompetenz. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Eds.). *Struktur, Entwicklung und Förderung von Lesekompetenz: Vertiefende Analysen im Rahmen von PISA 2000* (pp. 37-60). Wiesbaden: VS Verlag.

Karing, C, Lorenz, C., Matthäi, J., Pfost, M., Schurtz, I. M., Dörfler, T., & Artelt, C. (in prep.). *Skalenhandbuch BiKS-8-14*.

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

Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices*. New York, NY: Springer.

Köller, O., Baumert, J., & Schnabel, K. (2000). Zum Zusammenspiel von schulischem Interesse und Lernen im Fach Mathematik: Längsschnittanalysen in den Sekundarstufen I und II. In U. Schiefele & K.-P. Wild (Eds.), *Interesse und Lernmotivation: Untersuchungen zu Entwicklung, Förderung und Wirkung* (pp. 163-181). München: Waxmann.

Krapp, A. (2000). Interest and human development during adolescence: An educational-psychological approach. In J. Heckhausen (Ed.), *Motivational psychology of human development* (pp. 109-128). London: Elsevier.

Krapp, A. (2002). An educational-psychological theory of interest and its relation to self-determination theory. In E. Deci & R. Ryan (Eds.), *The handbook of self-determination research* (pp. 405-426). Rochester: University of Rochester Press.

Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learning and Instruction, 15*, 381-395. doi: 10.1016/j.learninstruc.2005.07.007

Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where these differences lie. *Journal of Reasearch in Reading, 32*, 199-214. doi: 10.1111/j.1467-9817.2008.01389.x

Lüftenegger, M., Schober, B., van de Schoot, R., Wagner, P., Finsterwald, M., & Spiel, C. (2012). Lifelong learning as a goal: Do autonomy and self-regulation in school result in well prepared pupils? *Learning and Instruction, 22*, 27-36. doi: 10.1016/j.learninstruc.2011.06.001

Marks, G. N. (2005). Accounting for immigrant non-immigrant differences in reading and mathematics in twenty countries. *Ethnic and Racial Studies, 28*, 925-946. doi: 10.1080/01419870500158943

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76* (2), 397-416.

McElvany, N., Kortenbruck, M., & Becker, M. (2008). Lesekompetenz und Lesemotivation: Entwicklung und Mediation des Zusammenhangs durch Leseverhalten. *Zeitschrift für Pädagogische Psychologie, 22*, 207-219. doi: 10.1024/1010-0652.22.34.207

McElvany, N., & Becker, M. (2010). Welche Prädiktoren braucht man zur Vorhersage von Lesekompetenz? Eine Kommunalitätsanalyse zur Bestimmung der uniquen und geteilten Varianzaufklärung psychologischer und soziologischer Konstrukte. In W. Bos, E. Klieme & O. Köller (Eds.) *Schulische Lerngelegenheiten und Kompetenzentwicklung: Festschrift für Jürgen Baumert* (pp. 147-172). Münster: Waxmann.

Meece, J. L., Glienke, B. B., & Burg, S. (2006). Gender and motivation. *Journal of School Psychology, 44*, 351-373. doi:10.1016/j.jsp.2006.04.004

Merkens, H., & Nauck, B. (1993). Ausländerkinder. In M. Markefka & B. Nauck (Eds.). *Handbuch der Kindheitsforschung* (pp. 447-457). Neuwied: Luchterhand.

Millard, E. (1997). Differently literate: Gender identity and the construction of the developing reader. *Gender and Education, 9* (1), 31-48. doi: 10.1080/09540259721439

Miller, S. D., & Faircloth, B. S. (2009). Motivation and reading comprehension. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 307-322). New York: Routledge.

Möller, J., & Schiefele, U. (2004). Motivationale Grundlagen der Lesekompetenz. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 101-124). Wiesbaden: VS Verlag.

Mullis, I. V.S., Martin, M. O., Foy, P., & Drucker, K. T. (2012). *PIRLS 2011 international results in reading.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Authors.

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

OECD. (2003). *Literacy Skills for the World of Tomorrow: Further Results from PISA 2000.* PISA: OECD Publishing.

Paratore, J. R., Cassano, C. M., & Schickedanz, J. A. (2011). Supporting early (and later) literacy development at home and at school: The long view. In M. Kamil, P. Pearson, E. B. Moje & P. Afflerbach (Eds.), *Handbook of reading research* (pp. 107-135). Vol. IV. New York: Routledge.

Paris, S.G., Lipson, M.Y., & Wixson, K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8,* 293-316. doi: 10.1016/0361-476X(83)90018-8

Pfost, M., Dörfler, T., & Artelt, C. (2010). Der Zusammenhang zwischen außerschulischem Lesen und Lesekompetenz: Ergebnisse einer Längsschnittstudie am Übergang von der Grund- in die weiterführende Schule. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42,* 167-176. doi: 10.1026/0049-8637/a000017

Philipp, M. (2008). *Lesen, wenn anderes und andere wichtiger warden: Empirische Erkundungen zur Lesesozialisation in der peer group bei Kindern aus fünften Klassen.* Münster: LIT Verlag.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Los Angeles/ London/ New Dehli/ Singapore: SAGE.

Renninger, K. A. (1992). Individual interest and development: Implications for theory and practice. In K. A. Renninger, S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 361-396). Hillsdale: Lawrence Erlbaum Associates.

Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone, & J. M. Harackiewicz (Eds.), Intrinsic and extrinsic motivation: the search for optimal motivation and performance (pp. 373-404). New York: Academic Press.

Renninger, K. A., Ewen, L., & Lasher, A. K. (2002). Individual interest as context in expository text and mathematical word problems. *Learning and Instruction, 12,* 467 – 491.

Retelsdorf, J., & Möller, J. (2008). Entwicklungen von Lesekompetenz und Lesemotivation: Schereneffekte in der Sekundarstufe? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 40*, 179-188. doi: 10.1026/0049-8637.40.4.179

Retelsdorf, J., Köller, O., & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and instruction, 21*, 550-559. doi: 10.1016/j.learninstruc.2010.11.001

Rheinberg, F. (1998). Theory of interest and research on motivation to learn. In L. Hoffmann, A. Krapp, K. A. Renninger & J. Baumert (Eds.), *Interest and learning* (pp. 126-145). Kiel: IPN.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55,* 68-78. doi: 10.1037110003-066X.55.1.68

Schaffner, E., Schiefele, U., & Schneider, W. (2004). Ein erweitertes Verständnis der Lesekompetenz: Die Ergebnisse des nationalen Ergänzungstests. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Eds.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 197-242). Wiesbaden: VS Verlag.

Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading, 3*, 257-279. doi: 10.1207/s1532799xssr0303_4

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 183-212). Hillsdale: Lawrence Erlbaum Associates.

Schiefele, U., & Schiefele, H. (1997). Motivationale Orientierungen und Prozesse des Wissenserwerbs. In H. Gruber & A. Renkl (Eds.), *Wege zum Können* (S. 14-31). Bern: Huber.

Schleicher, A. (2006). Where immigrant students succeed: A comparative review of performances and engagement in PISA 2003. *Intercultural Education, 17*, 507-516.

Stanat, P., Rauch, D., & Segeritz, M. (2010). Schulbezogene Motivation und Aspiration von Schülerinnen und Schülern mit Migrationshintergrund. 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. 200-230). Münster: Waxmann.

Torppa, M., Poikkeus, A.-M., Laakso, M.-L., Tolvanen, A., Leskinen, E., Leppänen, P.H.T., et al. (2007). Modeling the early paths of phonological awareness and factors supporting its development in children with and without familial risk of dyslexia. *Scientific Studies of Reading, 11*, 73-103, doi: 10.1080/10888430709336554.

Wigfield, A., & Asher, S. R. (1984). Social and motivational influences on reading. In P. D. Pearson, R. Barr, M. L. Kamil & P. Mosenthal (Eds.). *Handbook of reading research* (pp. 423-452). New York: Longman.

Wigfield, A., & Guthrie, J. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89* (3), 420-432.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A., Freedman-Doan, K., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A three-year study. *Journal of Educational Psychology, 89* (3), 451-569.

Wu, M. L., Adams, R. J., Wilson, M., & Haldane, S. A. (2007). *ACER ConQuest version 2.0: generalised item response modelling software.* Camberwell: ACER Press.

Chapter 8

8 Reading Literacy Development in Secondary School and the Effect of Differential Institutional Learning Environments

Maximilian Pfost and Cordula Artelt

Summary

The German secondary school system is characterized by a relatively early separation of students into different types of schools or school tracks that provide different types of curricula in accordance with the prerequisites of the learners. The stratification of the students into the different school tracks is based mainly on student achievement in elementary school, but is also influenced by other factors such as the socioeconomic status or immigration background of the family. As upper academic track schools should provide more favorable developmental conditions with regard to the students' cognitive competencies due to institutional characteristics and school composition effects, pre-existing differences in reading comprehension and vocabulary between the students in the different school tracks should further increase over the course of secondary school. In tracing the development of reading comprehension and vocabulary between Grade 5 and

Author Note

Maximilian Pfost,
Department of Educational Research, University of Bamberg, Germany.

Cordula Artelt,
Department of Educational Research, University of Bamberg, Germany.

This research was supported by grant WE 1478/4-1 & AR 301/9-1 from the German Research Foundation (DFG).

We would like to thank Benjamin Nagengast (University of Tübingen) for his insightful comments on a draft of this chapter.

Correspondence concerning this chapter should be addressed to Maximilian Pfost, Department of Educational Research, University of Bamberg, Markusplatz 3, 96045, Bamberg, Germany. E-mail: maximilian.pfost@uni-bamberg.de

Grade 7 in the current study, results indicated a widening gap between upper, middle, and lower academic track school students' reading comprehension, whereas stable achievement differences in vocabulary were found. A second analysis investigated the effect of attending the different school tracks while controlling for selectivity into the different secondary schools. Results indicated substantial positive effects of attending an upper academic track school in comparison to the lower and middle academic track schools in terms of effect sizes for reading comprehension and vocabulary, though not all results reached statistical significance. Taken together, favorable learning environments seem to support reading literacy development, but the reported findings should be generalized cautiously.

In most German states, students enroll in secondary school when they reach the age of 10 after 4 years of primary education (Cortina, Baumert, Leschinsky, Mayer, & Trommer, 2008; Faust, 2006). The secondary school system in Germany, in contrast to the primary education system, is marked by a strict institutional stratification of students into different types of schools or tracks that go along with distinct school leaving certificates and that provide different learning opportunities to their students. With regard to reading literacy, the transition from primary to secondary school is also marked by different conceptions of schooling and the function of reading. Whereas during primary school, instruction focuses on teaching children to read, over the course of secondary school, students increasingly read to learn (Burns & Kidd, 2010; Chall, 1983). Nevertheless, although explicit instruction in reading is rare and the process of acquiring further reading skills becomes increasingly incidental in the course of secondary school, there is still a generally positive trend in the development of students' reading literacy until students leave school (Hill, Bloom, Black, & Lipsey, 2008; Klicpera, Schabmann, & Gasteiger-Klicpera, 1993). Therefore, it is of critical importance to investigate the role of schools in a secondary school system that is characterized by an explicit between-school tracking for the development of reading literacy.

As mentioned, the German secondary school system separates their students by different types of schools or tracks that provide different types of curricula in accordance with the competencies and prerequisites of the learners. We call this form of organizational differentiation *between-school tracking* or *curricular differentiation by school type* (LeTendre, Hofer, & Shimizu, 2003) in contrast to forms of tracking that take place within schools (e.g., differentiating by courses or streams that can often be found in U.S. high schools). Thereby, the assignment of students to the different types of schools depends primarily on an interplay between decisions made by the primary schools and by the parents (Cortina & Trommer, 2005; Faust, 2005). Over the course of the last year in primary school, the school provides a recommendation for the educational career of the student. This recommendation is primarily based on the student's aptitudes, but also takes into account other prognostic factors (e.g., familial support of the child). The bindingness of this recommendation varies between the federal states, providing different scopes for parents' decision making with regard to the educational careers of their children. In the end, this procedure leads to a separation of the students between the different types of schools according to the students' cognitive abilities but also according to their social and familial backgrounds (Baumert & Köller, 2005; Baumert & Schümer, 2001; Ditton & Krüsken, 2006; Ditton, Krüsken, & Schauenberg, 2005). The rationale behind this institutional separation of students, which Gamoran and Mare (1989) call the *Positive View of Tracking*, is "that students differ in their academic goals and in the environments in which they learn best. Ideally, a system of academic tracking matches students' aptitudes with the objectives and learning environments to which they are best suited" (Gamoran & Mare, 1989, p. 1148). Therefore, a homogenization of the group of students with regard to their ability level should ideally enhance learning for all students (Baumert, 2006). Nevertheless, empirical support for this assumption has been mixed (cf. Ariga & Brunello, 2007; Slavin, 1990).

However, focusing exclusively on the question of the productivity of tracking practices in comparison to nontracking practices on students' learning neglects a second outcome dimension: individual differences or performance inequality between students who attend different tracks. Separating students into different school tracks might, for example, be very effective for students in higher academic tracks, whereas it might have detrimental effects for students in lower academic tracks. Of course, the opposite could also be true. Students in lower academic tracks might receive the instruction they need to catch up to the achievement level of the higher track students. Therefore, the following two questions require further analysis: How do the cognitive competencies of students who were separated into different academic tracks develop and how would these competencies have developed if the students who were assigned to a certain school track would have been assigned to another track?

Type of School and Causes of Individual Differences in Competence Development

In most German states, the secondary school system is comprised of at least three types of schools or tracks (Cortina, et al., 2008): a *lower academic track* (“Hauptschule”) that provides 5 years of basic secondary education, generally preparing students for vocational training; a *middle academic track* (“Realschule”), comprising 6 years of secondary education; and a *higher academic track* (“Gymnasium”) that comprises 8/9 years of secondary education and qualifies students for university admission. In addition, some German states run comprehensive secondary schools, offering all three types of school leaving certificates. As different types of schools pursue different academic goals and students are selected into these types of schools primarily according to their cognitive abilities and academic achievement, different learning environments are the result. These school-type-specific environments provide differential developmental possibilities for students based on differential distributional processes of economic, social, and cultural resources; differential institutional working and learning conditions; as well as differential school-type-specific educational and curricular traditions (Baumert, 2006; Baumert, Köller, & Schnabel, 1999; Baumert & Schümer, 2001; Gamoran & Berends, 1987). For example, whereas in lower academic track schools, it is still common to have a form teacher who teaches several or almost all subjects (Leschinsky, 2008a), teachers in middle or upper academic track schools are usually specialized to teach only two or three subjects (Leschinsky, 2008b; Trautwein & Neumann, 2008). In addition, upper academic track teachers tend to have higher levels of content knowledge as well as pedagogical content knowledge (Baumert, et al., 2010). Furthermore, comparing the cultures of instruction, relatively clear-cut differences between tracks are apparent: In the upper academic track schools, lessons are usually characterized by a high level of cognitive activation and a low level of teacher support, whereas in lower academic track schools, lessons are usually characterized by a high level of teacher support and a low level of cognitive activation (Kunter, et al., 2005). Finally, instruction in lower tracks often seems to proceed more slowly and is conceptually simplified, thereby providing only restricted access to knowledge for students who attend this track (Gamoran & Berends, 1987).

In addition to the thus-far described institutional differences in instruction, the student composition itself might support or handicap learning processes (Baumert, Stanat, & Watermann, 2006; Harker & Tymms, 2004; Pfost, 2011; Zimmer & Toma, 2000). This means that differences in the development of cognitive competencies might be attributable not only to institutional differences in the learning environments, but might also reflect differences in the characteristics of the students within these schools. For example, it has been shown that the proportion of students with an immigration background is negatively linked to the development of the students' reading competence (Pfost, 2011; Stanat, 2006; Walter & Stanat, 2008). Further studies have shown a positive relation between the mean level of achievement and individual reading development (Baumert, et al., 2006; Dreeben & Barr, 1988; Lehmann, 2006) or mathematics (Lehmann, 2006; Opdenakker, van Damme, de Fraine, van Landeghem, & Onghena, 2002; Zimmer & Toma, 2000). Finally, evidence exists for a positive effect of the aggregated mean socioeconomic status on students' academic achievement (Dumay & Dupriez, 2007; Ma & Klinger, 2000; van Ewijk & Sleegers, 2010). As the access to different school tracks is highly selective, institutional differences in the composition of students within schools is the result and may reinforce existing institutional differences in the learning opportunities that are offered. Consequently, different learning rates between students attending different school tracks in secondary school should be expected.

When reviewing differences in the development of cognitive competencies, a third cause of individual differences needs to be taken into account: differential learning rates due to individual characteristics or traits of the students themselves. Therefore, differences in competence development between different school tracks might be attributable to observed and unobserved characteristics that govern the selectivity of students into the different types of schools. A well-supported fact is that in primary school, students already differ in their school performances, familiar and social backgrounds, as well as expectations concerning future school achievement (Ditton & Krüsken, 2006; Gamoran & Mare, 1989; Maaz, Hausen, McElvany, & Baumert, 2006; Schneider & Stefanek, 2004). For example, parents from different economic and educational backgrounds might apply different strategies such as the utilization of paid private tutoring to realize their educational aspirations and therefore might try to actively influence the selection process into secondary school (Dang & Rogers, 2008; Schneider, 2004). Furthermore, students differ in their prior knowledge when entering secondary school, which might directly result in different learning rates (Renkl, 1996). Within the domain of reading, Stanovich (1986, 2000) describes a model of increasing interindividual differences in reading literacy; he named this the Matthew effect model. Thereby, the cumulative advantages of good readers or the cumulative disadvantages of bad readers are the result of reciprocal self-reinforcing causal processes: “The very children who are reading well and who have good vocabularies will read more, learn more word meanings, and hence read even better. Children with inadequate vocabularies – who read slowly and without enjoyment – read less, and as a result have slower development of vocabulary knowledge, which inhibits further growth in reading ability” (Stanovich, 1986, p. 381). However, empirical studies that have investigated the Matthew effect model in reading have produced mixed results. On the one hand, there is much empirical support from longitudinal studies concerning the reciprocal relation of reading ability, reading motivation, and reading behavior (McElvany, Kortenbruck, & Becker, 2008; Morgan & Fuchs, 2007; Pfost, Dörfler, & Artelt, 2010). On the other hand, studies that have focused on the development of the competence gap between good and poor readers have not yet accumulated convincing evidence which clearly supports a pattern of increasing or a pattern of decreasing differences in reading achievement over time (e.g. Aarnoutse, van Leeuwe, Voeten, & Oud, 2001; Bast & Reitsma, 1998; Kempe, Eriksson-Gustavsson, & Samuelsson, 2011; Parrila, Aunola, Leskinen, Nurmi, & Kirby, 2005; Pfost, Dörfler, & Artelt, 2012).

In sum, differences in learning rates between students attending lower, middle, and upper academic track schools are the result of an interplay between individual, institutional, and school composition factors that may add up, reinforce, or compensate each other over the course of students' individual development.

Achievement Differences and Achievement Growth in Secondary School – Empirical Findings

Cross-sectional studies, especially the four PISA studies run by the OECD between 2000 and 2009 (Baumert, et al., 2001; Klieme, et al., 2010; Prenzel, et al., 2007; Prenzel, et al., 2005), have reported large differences in cognitive competencies between the students who attend different school tracks in Germany. In the most recent PISA study, 15-year-old students attending upper academic track schools on average achieved a reading comprehension score that was more than one and a half standard deviations above the average reading comprehension score of students attending lower academic track schools. Students attending middle academic track schools as well as comprehensive schools reached an average reading comprehension score in between these other two types of schools (Naumann, Artelt, Schneider, & Stanat, 2010). Comparable results have been reported for mathematics and science (Frey, Heinze, Mildner, Hochweber, & Asseburg, 2010; Rönnebeck, Schöps, Prenzel, Mildner, & Hochweber, 2010). Intuitively, we might conclude that these differences are the result of achievement differences prior to secondary school plus different learning rates between school tracks, but cross-sectional studies such as PISA cannot determine the time in the course of development at which differential learning rates appear. Thus, the hypothesis of a widening achievement gap between the different academic tracks needs to be analyzed longitudinally.

Within the domain of mathematics, the assumption of a widening achievement gap has been investigated and verified several times (Becker, Lüdtke, Trautwein, & Baumert, 2006; Köller & Baumert, 2001) with the exception of Schneider and Stefanek (2004), who reported stable mathematics achievement differences between Grade 2 and Grade 11. The reported results from Germany converge well with studies that have investigated the effect of taking advanced courses in U.S. high schools (Gamoran & Mare, 1989; Schmidt, 2009).

Within the domain of reading, however, studies have been less frequent and the results have been more controversial. This might, at least partially, be attributable to differences in the learning opportunities that underlie the development of different cognitive skills (cf., Köller & Baumert, 2008). Whereas for the development of mathematical skills, schools play almost a monopolistic role in the transfer of knowledge, within the domain of reading, further learning opportunities such as leisure time reading (e.g., Pfost, Dörfler, et al., 2010; Spear-Swerling, Brucker, & Alfano, 2010) are of high relevance. Consequently, it might be reasonable to expect that differences in school learning environments might be more related to the development of mathematics than to the development of reading literacy. Retelsdorf and Möller (2008), in analyzing data from the LISA study, reported small but nonsignificant differences in the development of reading literacy from Grade 5 to Grade 6 between lower (d = 0.59), middle (d = 0.62), and upper academic track schools (d = 0.82). Initial differences in reading literacy in Grade 5, when students enter secondary school, however, were already relatively large, with students in the upper academic track scoring on average more than one standard deviation (d = 1.22) above students from the middle academic track and even more than two standard deviations (d = 2.30) above students from the lower academic track. Similar results were presented by Gröhlich, Bonsen, and Bos (2009): In analyzing data from more than 10,000 students from the Hamburg KESS study, the authors reported the highest growth in reading literacy between the end of Grade 4 and Grade 6 for students who attended comprehensive schools (d = 0.47), followed by students who attended lower and middle academic track schools (d = 0.45). The lowest average growth was reported for upper academic track students (d = 0.42). The results confirm the findings from the antecedent LAU study (Lehmann, Peek, Gänsfuß, & Hußfeldt, 1998). Taken together, the results in the domain of reading have been less stringent and have not confirmed the assumption of a widening gap over the course of secondary school.

The question of whether a privileged school learning environment is linked to an increased learning rate was also addressed by the Berlin ELEMENT study (Lehmann & Lenkeit, 2008), which was subsequently reanalyzed by Baumert, Becker, Neumann, and Nikoleva (2009). In the state of Berlin, students have the opportunity to switch to some upper academic track schools ("grundständiges Gymnasium") after Grade 4 or to stay in a prolonged elementary school and change to secondary school after Grade 6. Students who chose to attend early upper academic track schools after Grade 4 had, in comparison to the students who remained in elementary school, better marks, better reading, and mathematics competencies and came from families with a higher socioeconomic status. Results describing the competence development between Grade 4 and Grade 6 showed, beyond initial differences in reading literacy, a comparable learning rate for students in the two types of schools. With regard to mathematics, students in the early upper academic track school showed an increased learning rate in comparison to the elementary school students. The reanalysis of the data by Baumert et al. (2009), however, focusing on the role of the learning environment on the development of reading and mathematics, did not demonstrate a more favorable learning rate in reading or in mathematics for students in the early upper academic track schools after students' individual characteristics, driving the transition from elementary to early upper academic track school, had been taken into account. Therefore, the hypothesis that a privileged learning environment leads to higher learning rates was not confirmed by this study. Finally, using data from the BiKS study, Pfost, Karing, Lorenz, and Artelt (2010) report a widening achievement gap or fan-spread effect between students attending the lower academic track and the middle as well as upper academic track for reading comprehension, but not vocabulary, between Grade 5 and Grade 6. In addition, a fan-spread effect between students attending different secondary schools was already traceable when students still attended primary school.

Taken together, whereas in the domain of mathematics, fan-spread effects have been demonstrated several times, within the domain of reading, results have been less stringent and have mostly indicated relatively stable achievement differences between different types of schools across the course of secondary school. However, due to the assumption of different learning environments, also fan-spread effects in the domain of reading can be expected.

Research Questions

The current study focused on the following two questions: First, can differences in the development of reading literacy by type of school/school track be found? With regard to the assumption that upper academic track schools provide a favorable learning environment due to institutional and compositional factors and that students attending upper academic track schools on average have higher cognitive abilities, which should additionally promote further learning, different learning rates in favor of students in upper academic track schools were expected. Furthermore, as lower academic track schools should provide the least favorable learning conditions, the lowest learning rates were expected within this school type. Second, it seemed important to ask whether an effect of attending different types of schools on reading achievement measures could be verified independent of students' characteristics that govern the selectivity into the different secondary school tracks. Again, we expected a favorable effect of attending upper academic track schools in comparison to middle and lower academic track schools, after controlling for important covariates that go along with the choice of a certain track. Due to sample-size restrictions, students from middle and lower academic track schools were grouped together. Therefore, only the effect of attending upper academic track schools in comparison to attending an alternative type of school (middle and lower academic tracks) was estimated.

The current paper extends the findings reported by Pfost, et al. (2010) in at least two ways: at first, data up to Grade 7 was available. Second, the role of covariate selection for the estimation of effects of different institutional learning environments was addressed in more detail.

Method

Design and Participants

All analyses were based on data from the BiKS-8-14 panel study. At the first point of measurement, in the second term of Grade 3, $N = 2{,}395$ students were assessed. After the transition from primary into secondary school, a subsample of $n = 922$ students (38.5% of the original sample) was further followed across secondary school ($n = 268$ in the lower, 188 in the middle, and 466 in the upper academic tracks). Students were selected for further participation in the BiKS-8-14 panel study when they agreed to participate further, when they chose a school within the BiKS inquiry region that had at least one class with at least three participants, and when the school was not characterized by comprehensive or remedial instruction (cf., Schmidt, Schmitt, & Smidt, 2009). Furthermore, n = 879 secondary school students (n = 102 in the lower, 135 in the middle, and 642 in the upper academic tracks) were additionally recruited in Grade 5 for participation in the BiKS panel study, resulting in a total sample of N = 1,801 secondary school students. Whereas in primary school, data collection took place every half year (Measurement Waves 1, 2, and 3), in secondary school, data were collected annually at the end of each academic year (Measurement Waves 4, 5, and 6). The following analyses focused on the development of measures of reading comprehension and vocabulary between Grade 5 and Grade 7. Additional data from the elementary school years were taken into account for the second set of analyses. The average age of the students was 11.4 years (SD = 0.5) in Grade 5. Furthermore, in our sample, 13.8% of the students lived in households with immigration backgrounds. The gender of the students was almost equally distributed; 47.8% of the students were male and 52.2% were female.

Measures

Students, teachers, and parents were tested on a wide range of measures. In the following section, the measures that were used in the current analysis are presented. At first, the two measures of reading comprehension and vocabulary used in secondary school (Grade 5 to 7) are depicted. Developmental differences between school tracks on these two variables are of major interest in our analyses. Therefore, these two variables are presented in detail. Subsequently, the variables/covariates that were used in the second analysis, in order to control for the selectivity into the different school tracks, are depicted. All covariates were assessed in primary school.

Reading comprehension. In Grade 5, reading comprehension was measured by a sample of six short texts with a total of 43 multiple-choice items developed by the BiKS research group. For the reading comprehension test, the students had to read a given text, search relevant information, and generate more or less high inferences from the text to answer the given items. In Grade 6, three texts with a total of 31 multiple-choice items were used. Finally, in Grade 7, again, three texts with a total of 26 multiple-choice items were used. For the three waves of measurement, a common item design with a nonequivalent groups/anchor-item test design was applied (Holland, Dorans, &

Peterson, 2007; Kolen & Brennan, 2004), allowing the estimation of students' reading comprehension on a common metric within an IRT framework. Therefore, for all reading comprehension test items, the item difficulty parameters were estimated with a three-dimensional 1-parameter Rasch model by using the ConQuest software package (Wu, Adams, Wilson, & Haldane, 2007). A design matrix was specified and the item difficulty parameters of the three waves of measurement were estimated in a single simultaneous run (concurrent estimation). Item difficulty parameters for the same items across different waves of measurement were set equal. Subsequently, individual students' abilities were estimated in a second run by weighted likelihood estimates (WLEs) for every wave of measurement using the item difficulty parameters of the concurrent estimation. Missing responses were treated as incorrect during the item calibration stage as well as during the estimation of the person parameters. The estimated individual ability scores were conclusively T-standardized (M = 50, SD = 10) in Grade 5. The reliabilities (WLE-reliability) of the reading comprehension measures were satisfactory for all waves of measurement ($\text{Reliability}_{\text{Grade 5}}$ = .78, $\text{Reliability}_{\text{Grade 6}}$ = .77, $\text{Reliability}_{\text{Grade 7}}$ = .76).

Vocabulary. Students' vocabulary was measured by a set of 35 items from the subscale V1 (Vocabulary) of the *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision* (KFT 4-12 + R; Heller & Perleth, 2000). Additional vocabulary items that were used in Grade 7 were disregarded in the present analysis in order to keep the metric constant. Ceiling effects were negligible as still in Grade 7 the maximum test score was reached by just one student of the sample. For every item, a target word as well as a selection of four additional words was presented for reading. Students had to indicate the word whose definition best matched the presented target word. Students' vocabulary was estimated by summing the number of correct answers. For ease of interpretation, students' vocabulary scores were also T-standardized (M = 50, SD = 10) in Grade 5 by a linear transformation. The internal consistency (Cronbach's α) of the vocabulary test was satisfactory for the three waves of measurement ($\alpha_{\text{Grade 5}}$ = .78, $\alpha_{\text{Grade 6}}$ = .80, $\alpha_{\text{Grade 7}}$ = .78).

Covariates. Socioeconomic and ethnic-cultural backgrounds. Data concerning students' socioeconomic and ethnic-cultural backgrounds were collected in a highly standardized telephone interview in the first and third waves of measurement in Grade

3 and Grade 4 of elementary school. In order to determine students' immigration backgrounds, parents were asked questions concerning their cultural origin. Students were classified as having an immigration background when at least one parent was born in a foreign country. Furthermore, the parents were asked questions concerning their familial, educational, as well as occupational status. With this information, the highest ISEI (International Socio-economic Index of Occupational Status; Ganzeboom, De Graaf, & Treiman, 1992) and educational level of the parents was determined.

Cultural capital. Parents were asked to specify the number of books they had at home. The responses were categorized by the interviewers. Categories ranged from 1 (*not one*) to 7 (*more than 500*).

Extracurricular reading behavior. Students' habitual extracurricular reading behavior was assessed by a single item ("Does [the name of the child] read for pleasure?") in the parental telephone interview in Grade 4. Parents rated the frequency of their children's reading behavior on a 4-point Likert-type scale with the response options 1 (*almost never or never*), 2 (*rarely*), 3 (*yes, several times a week*), and 4 (*yes, everyday*).

Reading self-concept. Students' reading self-concept was assessed by a single item ("How good are you in school in... reading?") in the students' questionnaire in Grade 4. Students rated their reading self-concept on a 4-point Likert-type scale ranging from 1 (*bad*) to 4 (*very good*).

Vocabulary. In Grade 4, students' vocabulary was measured by a set of 30 items from the supplementary vocabulary test of the culture fair intelligence test (CFT 20, german version: Weiß, 1987).

Mathematics competence. Students' mathematics competence in Grade 4 was measured by a selection of 19 items from the DEMAT 4 (Gölitz, Roick, & Hasselhorn, 2005).

Spelling. Spelling was measured in Grade 4 by using 21 items from the DRT 4 (Grund, Haug, & Naumann, 2003).

General cognitive abilities. Students' general cognitive abilities were assessed in Grade 4 with a set of 15 items from the matrices subtest of the culture fair intelligence test (CFT 20-R, german version: Weiß, 2006).

Reading comprehension. In Grade 4, reading comprehension was measured by a sample of 13 short texts with 20 multiple-choice items from the subscale text comprehension of the ELFE 1-6 (Lenhard & Schneider, 2005). The test was prolonged by adding three new texts with six multiple-choice items developed by the authors to avoid ceiling effects.

Grades. Information concerning the students' grades after the first term of Grade 4 was provided by the class teachers. In Germany, grades range from 1 (*excellent*) to 6 (*insufficient*).

Analytic Strategy

The first set of analyses addressed the question of whether differences in the development of reading comprehension and vocabulary between students attending different types of schools could be demonstrated. In order to test for developmental differences, difference scores for reading comprehension and vocabulary, using models of true intraindividual change (cf. Geiser, 2010; Steyer, Eid, & Schwenkmezger, 1997), were computed (Figure 1). The type of school was used as a grouping variable. As there was only one indicator of reading comprehension or vocabulary available for each wave of measurement, a latent achievement indicator was not estimated. Consequently, the measurement error of the manifest variables was set to zero. The initial unconstrained model was just identified, fitting the data perfectly. To test for differences between groups, mean change scores between different types of schools were set to be equal and compared to the model without this constraint. All multigroup models of difference scores were estimated with M*plus* 6.1 (Muthén & Muthén, 1998-2010). In order to take the nested data structure into account, the *type is complex* option was used. Although an MLR estimator was used, the chi-square value for testing the constrained model against the alternative, unconstrained (just-identified) model was not corrected as there was not yet a routine within M*plus* for doing this when missing data were replaced by multiple imputation.[1] The analyses were run two times. In the first analysis, students were grouped according to the type of school that these students attended in Grade 5. Changes in the school type between Grade 5 and Grade 7 that

[1] cf. M*plus* Discussion board, posting by Linda K. Muthén on 16th June 2006 on http://www.statmodel.com/discussion/messages/22/381.html [17th March 2012].

may have occurred were ignored. In the same way, students who had to repeat a class between Grade 5 and Grade 7 were treated as though they had advanced in the normal manner. In these two cases (change of school and grade repetition), test information of Grade 6 and/or Grade 7 was almost never available, and the achievement scores were imputed. To support the interpretation of our results, the models describing differences in reading comprehension and vocabulary development were reanalyzed in a second set of analyses, considering only students who were still actively participating in the study in Grade 7, who did not change their type of school, and who did not repeat a class during the time period under investigation.

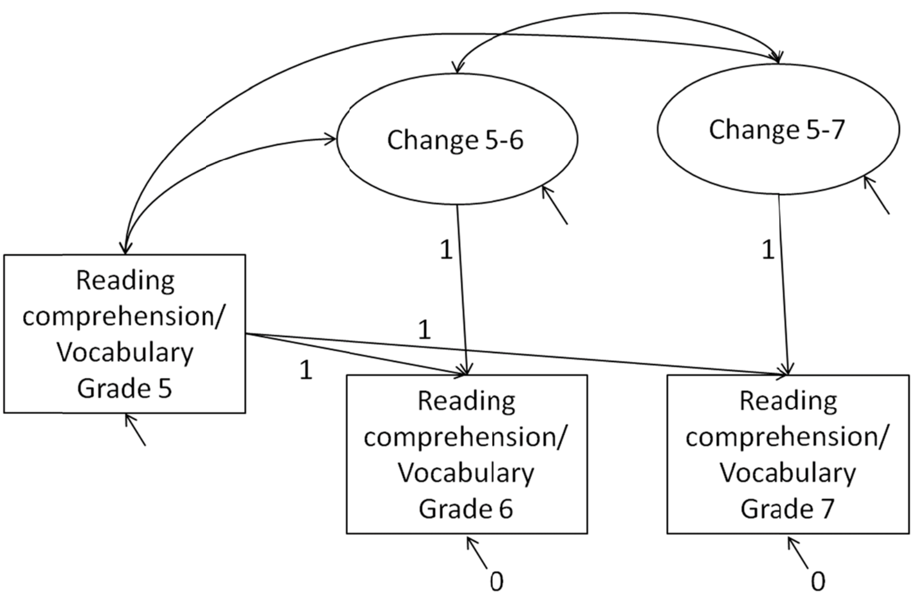

Figure 1. Specified difference score model for reading comprehension/vocabulary. The model is specified as baseline model.

The second set of analyses addressed the question of whether an effect of attending different types of schools or school tracks on the development of reading comprehension and vocabulary could be verified independent of individual characteristics influencing the selectivity into the secondary school system. To address this research question, we used the reduced subsample of students in secondary school for whom test information, including inter alia measures of reading comprehension and vocabulary from the elementary school years, was available. In order to disentangle institutional from individual effects, interindividual differences between students prior to their secondary school attendance needed to be adequately controlled. One of the most efficient tools for estimating treatment effects (e.g., the effect of attending different types of schools) in nonexperimental studies is Propensity-Score-Matching (PSM). In general, matching methods within observational studies aim to equate a distribution of covariates in treatment and control groups by drawing students from both groups who are similar on a set of observed covariates (Rosenbaum & Rubin, 1985; Stuart, 2010). Matching methods often come into operation when causal inferences about treatment effects in observational designs are of particular interest (c.f. Morgan & Winship, 2007; Rubin, 1997; West & Thoemmes, 2010). PSM traditionally comprises two analytical steps: First, for every student, the probability of being in either the treatment (TG) or the control group (CG) is calculated on the basis of the covariates that are taken into account. In the present analysis, attending an upper academic track school comprised the treatment condition and lower or middle academic track schools the control condition. In the current analysis, the following covariates were considered: the state where the school was located (dummy coded: 0 = Hesse, 1 = Bavaria), students' age and sex (dummy coded: 0 = female, 1 = male), parents' education (dummy coded: 0 = parents did not reach university entrance qualification, 1 = parents reached university entrance qualification), students' immigration background (dummy coded: 0 = no immigration background, 1 = students have an immigration background), parents' HISEI, cultural capital of the parents (the categories were dummy coded), students' time spent in extracurricular reading (the categories were dummy coded), students' reading self-concept (the categories were dummy coded), and Grade 4 achievement measures of vocabulary, mathematics, spelling, general cognitive abilities, and reading comprehension. Only linear effects of the covariates were considered. In the second matching analysis, in addition to the already denoted variables, students' grades after the first term of Grade 4 in mathematics and German were taken into account. As denoted, students' grades from the first term of Grade 4 were directly linked to the choice of school track. However, school grades are often not comparable to each other due to different applied reference scales (Maaz, et al., 2008; Trautwein, Lüdtke, Becker, Neumann, & Nagy, 2008; Treutlein & Schöler, 2009) and should therefore be treated and interpreted with caution.

On the basis of these variables, a probit score which indicates a student's probability of attending the upper academic track school (TG) given that student's covariates was estimated. Then, students in the two groups were matched to each other on the basis of the calculated probit score using radius matching (see Dehejia & Wahba, 2002; Morgan & Winship, 2007). Therefore, for each treatment case control cases were selected that were located within a particular distance – the radius – of the calculated propensity score. In cases in which more than one control student was located within the maximum acceptable distance around the treatment group student, the selected control cases were given equal weights. The radius was set at $\delta = 0.005$. Treatment cases that did not have a possible counterpart within the control cases were said to be off the support and were not considered for further analysis. The same was true for control cases without possible counterparts from the treatment cases. Therefore, the interpretability of the treatment effect was limited to those for whom possible counterparts existed (common-support treatment effect for the treated). In other words, the estimated average effect of attending an upper academic track school (TG), in comparison to attending lower or middle academic track schools (CG), on the development of reading comprehension and vocabulary is only informative with regard to those students who typically attend an upper academic track school and for whom comparable counterparts who attend lower and middle academic track schools exist. As mentioned, students attending lower and middle academic track schools were grouped together because of their small sample size. After the matching procedure, balance with respect to the incorporated covariates and the overlap between the two groups was checked. Therefore, the standardized differences of the covariates between the two treatment groups before and after the matching procedure were computed. In the final step, the analysis of the outcomes, differences in reading comprehension and vocabulary in Grade 7 between the matched groups were tested. Propensity-Score-Matching was done with STATA 11 using the psmatch2 routine (Leuven & Sianesi, 2003).

Missing data. Missing data is a typical problem of research in the social sciences, especially in longitudinal studies. In the current study, missing data may have occurred on the one hand because parents did not give consent for their child to participate in the study. What is known from the literature is that active informed parental consent is related to factors such as the degree of deviant behavior of the students, students' scholastic performance, and the social and ethnic backgrounds of families (Courser, Shamblen, Lavrakas, Collins, & Ditterline, 2009; Esbensen, et al., 1996; Esbensen, Hughes Miller, Taylor, He, & Freng, 1999; Unger, et al., 2004). On the other hand, parents may have given their informed consent but students might not have been present on the testing day, might not have correctly answered the questions, or may have left the study after a certain amount of participation (dropout). Study dropout in particular may be a sign of educational problems such as repeating a year or changing school type, and therefore needs to be treated cautiously (van de Grift, 2009). In other words, treatment-related attrition may be a serious threat to the internal validity of the estimated results (West & Thoemmes, 2010). In the first analysis, the data of all secondary school students in schools in which competence measurement took place and for whom parental consent was present were included in the analysis. Missing data on measures of reading comprehension and vocabulary were replaced by multiple imputation (m = 5) using a broad set of auxiliary variables. Multiple imputation was implemented by using an R script by Robitzsch (personal communication, March 18, 2011) controlling the imputation with Partial Least Squares regression within MICE (van Buuren & Oudshoorn, 2000). In order to verify the results of the first descriptive analysis, a second descriptive analysis was run by which, again, a dataset to which multiple imputation was applied was used, but the analysis was restricted to students who were still actively participating in the study in Grade 7, who did not change their type of school, and who did not repeat a class during the time period under investigation. We will denote this reduced sample as the "active sample" as students were still actively participating in the study in Grade 7. Finally, an EM algorithm that applied single imputation was used on the covariates that were used in the Propensity-Score-Matching. Although single imputation does not seem to be an adequate strategy in outcome analyses, it seems to be a sufficient and effective approach in the context of Propensity-Score-Matching (Stuart, 2010). The propensity score matching analysis was run exclusively using the active subsample of n = 658 students, for whom data from the primary school years were available and who were still active participants in the BiKS-8-14 longitudinal study in Grade 7.

Results

Developmental Differences in Reading Comprehension and Vocabulary

In order to trace interindividual differences in the development of reading comprehension and vocabulary, difference scores based on models of true intraindividual change were computed. The models were specified as baseline models, allowing for the analysis of differences in changes in reading comprehension and vocabulary between Grade 5 and Grade 6 (Change 6-5) as well as Grade 5 and Grade 7 (Change 7-5). A graphical illustration of the development of reading comprehension and vocabulary by type of school for the entire sample of secondary school students is depicted in Figures 2 and 3. The corresponding estimated results are presented in Table 1.

Table 1. Reading Comprehension and Vocabulary Development by School Track

	Grade 5 *M (SD)*	Grade 6 *M (SD)*	Grade 7 *M (SD)*	Change 5-6 *M (SD)*	Change 5-7 *M (SD)*
	Reading comprehension				
Lower academic track	40.47 (8.47)	41.98 (9.16)	43.80 (11.31)	1.51 (10.30)	3.33 (11.25)
Middle academic track	47.60 (7.77)	50.49 (9.41)	50.93 (11.80)	2.90 (8.88)	3.34 (11.26)
Upper academic track	53.90 (8.58)	58.21 (11.36)	60.26 (13.97)	4.32 (10.61)	6.36 (12.83)
Full sample	50.01 (10.00)	53.49 (12.45)	55.20 (14.74)	3.49 (10.32)	5.20 (12.34)
Test of significance[a]	$p < .01$[b]			$p < .01$	$p < .01$
	Vocabulary				
Lower academic track	40.84 (8.81)	45.13 (9.98)	50.22 (8.83)	4.29 (8.65)	9.38 (8.96)
Middle academic track	47.03 (7.92)	52.20 (9.53)	54.93 (9.10)	5.16 (8.27)	7.89 (8.95)
Upper academic track	53.92 (8.50)	58.54 (8.20)	61.09 (7.29)	4.62 (7.47)	7.17 (8.15)
Full sample	50.00 (10.00)	54.65 (10.35)	57.75 (9.14)	4.65 (7.88)	7.75 (8.52)
Test of significance[a]	$p < .01$[b]			*ns*	$p < .01$

Note. Sample size was $n = 370$ students in lower academic track schools, $n = 323$ in middle academic track schools, and $n = 1{,}108$ students in upper academic track schools.
[a]It was tested whether estimates were equal between students attending lower, middle and upper academic track schools.
[b]Mplus Type is General was used as Grade 5 reading comprehension/vocabulary was treated as manifest.

First, results indicated large differences in reading comprehension in Grade 5 between students in the different school tracks. Students attending upper academic track schools on average achieved the highest reading comprehension score, whereas students in the lower academic track schools achieved the lowest. Furthermore, significant differences in the development of reading comprehension between different school tracks were found: Between Grade 5 and Grade 6, students in the upper academic track schools showed the largest increase in reading comprehension, followed by students attending middle academic track schools. The smallest increase was measured in the group of lower academic track students.[2] A model constraint representing equal average reading comprehension development between the three type of schools was significant ($\Delta\chi^2$ = 12.212, *df* = 2, p < .01), indicating that developmental differences between school tracks are of statistical relevance. Regarding the development of reading comprehension for the full 2-year period between Grade 5 and Grade 7, we still found a clear statistically significant difference between students in the different school tracks ($\Delta\chi^2$ = 22.458, *df* = 2, p < .01). Again, students attending upper academic track schools showed the highest learning rate in comparison to lower and middle academic track students. The average learning rate of students attending lower academic track schools was comparable in size to the learning rate of the middle academic track students.

[2] Due to the application of a different scaling and imputation procedure as well as the usage of different analytic models, the reported growth rates may slightly vary from the results reported by Pfost, Karing, Lorenz, and Artelt (2010).

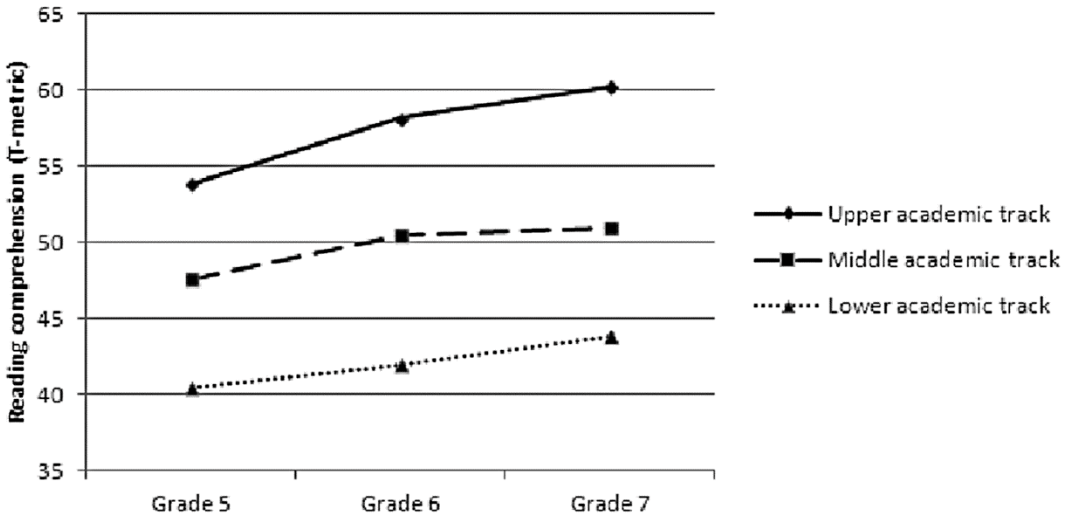

Figure 2. Development of reading comprehension by type of school. Estimates are based on the full sample of secondary school students (cf. Table 1 for corresponding data).

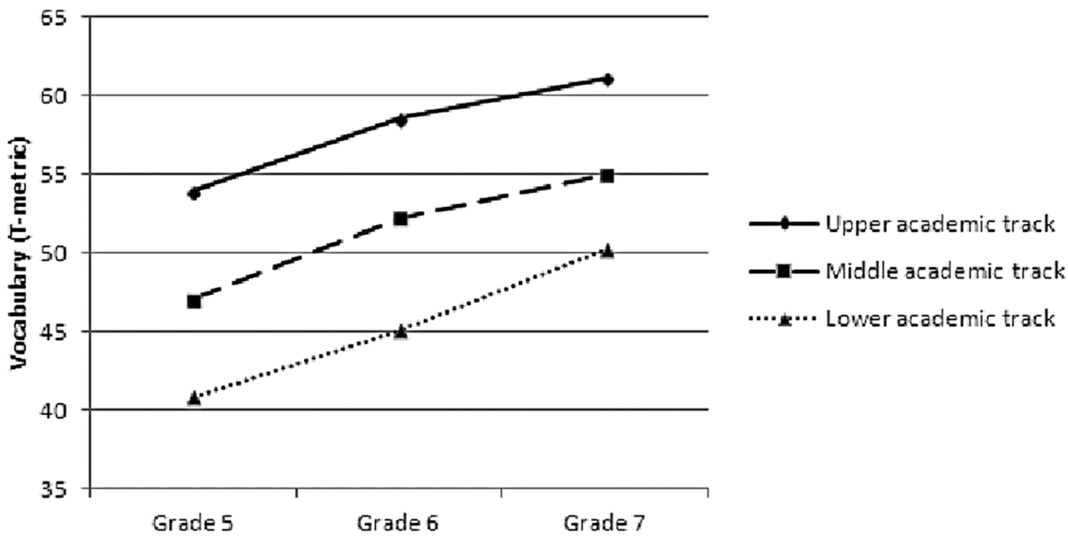

Figure 3. Development of vocabulary by type of school. Estimates are based on the full sample of secondary school students (cf. Table 1 for corresponding data).

Regarding vocabulary, again, strong interindividual differences in Grade 5 between students attending the different types of schools were present. When tracing the development of vocabulary between Grade 5 and Grade 6, no differences in the learning rate between students attending different types of schools were found ($\Delta\chi^2 = 1.220$, $df = 2$, *ns*). However, when analyzing the long-term development of vocabulary between Grade 5 and Grade 7, significant differences occurred ($\Delta\chi^2 = 10.144$, $df = 2$, $p < .01$). Interestingly, the developmental pattern was different from the one found for reading comprehension. Whereas for reading comprehension, the highest learning rate was found for students attending upper academic track schools; for vocabulary, the highest learning rate was found for students attending lower academic track schools. This means that lower academic track students caught up to the performance of the better performing middle and upper academic track students who were comparable in their learning rates.

In summary, results based on the full sample of secondary school students provide evidence for a widening gap or fan-spread effect for reading comprehension between students attending different school tracks, whereas with regard to the development of vocabulary, the opposite seems true: On average, students attending lower academic track schools showed the largest gains in vocabulary, whereas the smallest gains were found for upper academic track students.

Then, the same two difference score models for reading comprehension and vocabulary were estimated, but analyses were restricted to the sample of students who were still actively participating in the BiKS study in Grade 7, who did not change their type of school, and who did not have to repeat a class. This restriction reduced the sample size by $n = 443$ (24.6%) students, leading to an effective sample size of $n = 1{,}358$ (75.4% of the full sample) students. The reduced or active sample was composed of $n = 196$ (formerly $n = 370$; 53.0%) lower academic track students, $n = 267$ (formerly $n = 323$; 82.7%) middle academic track students, and $n = 895$ (formerly $n = 1{,}108$, 80.8%) upper academic track students. The estimated model results for the active sample are presented in Table 2.

Table 2. Reading Comprehension and Vocabulary Development by School Track (Active Sample)

	Grade 5 *M (SD)*	Grade 6 *M (SD)*	Grade 7 *M (SD)*	Change 5-6 *M (SD)*	Change 5-7 *M (SD)*
	Reading comprehension				
Lower academic track	40.47 (8.70)	42.37 (8.61)	43.75 (11.42)	1.90 (9.79)	3.28 (10.96)
Middle academic track	48.06 (7.57)	50.94 (9.09)	51.61 (11.65)	2.88 (8.78)	3.56 (11.38)
Upper academic track	54.57 (8.52)	59.51 (11.12)	61.76 (13.69)	4.94 (10.55)	7.19 (12.76)
Full sample	51.25 (9.80)	55.35 (12.17)	57.17 (14.66)	4.10 (10.19)	5.91 (12.38)
Test of significance [a]	$p < .01$[b]			$p < .01$	$p < .01$
	Vocabulary				
Lower academic track	40.87 (8.78)	45.33 (10.08)	49.67 (8.95)	4.46 (8.50)	8.81 (7.92)
Middle academic track	47.53 (7.76)	52.67 (9.15)	55.30 (9.10)	5.14 (8.10)	7.77 (8.67)
Upper academic track	54.85 (7.99)	59.65 (7.59)	62.06 (6.74)	4.80 (7.21)	7.21 (7.68)
Full sample	51.39 (9.58)	56.21 (9.82)	58.94 (8.90)	4.82 (7.59)	7.55 (7.94)
Test of significanc [a] e	$p < .01$[b]			*ns*	*ns*

Note. The estimates refer to students who were still actively participating in the BiKS study in Grade 7, who did not change their type of school, and who had not repeated a class during the time period under investigation (active sample). Sample size was n = 196 students in lower academic track schools, n = 267 in middle academic track schools, and n = 895 students in upper academic track schools.
[a]It was tested whether estimates were equal between students attending lower, middle and upper academic track schools.
[b]Mplus Type is General was used as Grade 5 reading comprehension/vocabulary was treated as manifest.

In comparison to the estimated results for the full sample (cf. Table 1), the estimations for the active sample (cf. Table 2) differed in two ways: First, the overall reading comprehension and vocabulary levels were about one tenth of a standard deviation higher in the reduced, active sample than in the full sample. This may be due to two causes. On the one hand, dropout was higher in lower academic track schools than in middle and upper academic track schools. On the other hand, especially within the upper academic track schools, students with lower achievement levels tended to drop out more often. Second, whereas in the first set of analyses, significant differences in the development of vocabulary between Grade 5 and Grade 7 between school tracks were found, analyses based on the active sample did not confirm this result ($\Delta\chi^2 = 3.543$, $df = 2$, *ns*). This difference might be attributable at least in part to a lower estimated vocabulary gain between Grade 5 and Grade 7 for students attending lower academic track schools in the active sample in comparison to the complete sample that included student dropouts. With regard to the development of reading comprehension, significant developmental differences in favor of students attending upper academic track schools were found, confirming the results of the first analysis that was based on the data of all secondary school students.

The Effect of Institutional Differences in Learning Environment on the Development of Reading Comprehension and Vocabulary

In order to test whether differences in the development of reading comprehension and vocabulary could be attributed to institutional differences in the learning environment, the selectivity of the students into the different school types had to be taken into account. Analyses were restricted to a subsample of n = 658 students, for whom information – inter alia test data – from the elementary school years was available and who were still active study participants in Grade 7 (active subsample). The developmental trends for reading comprehension and vocabulary for this longitudinal subsample of active secondary school students were comparable to the developmental trends for the full sample of active secondary school students (the full sample comprised also students that were not tested in primary school; cf. Tables 2 and 3).

Table 3. Reading Comprehension and Vocabulary Development by School Track (Active Elementary-Secondary-School Longitudinal Subsample)

	Grade 5 *M (SD)*	Grade 6 *M (SD)*	Grade 7 *M (SD)*	Change 5-6 *M (SD)*	Change 5-7 *M (SD)*
	Reading comprehension				
Lower academic track	40.27 (8.92)	42.20 (8.71)	42.76 (10.98)	1.92 (10.16)	2.48 (10.85)
Middle academic track	47.10 (7.42)	50.29 (9.47)	50.60 (12.12)	3.19 (9.19)	3.50 (11.67)
Upper academic track	53.71 (8.39)	58.13 (10.88)	61.34 (13.64)	4.43 (10.52)	7.64 (13.28)
Full sample	49.42 (9.88)	53.05 (11.99)	55.05 (14.89)	3.63 (10.21)	5.63 (12.67)
Test of significance [a]	$p < .01^{b}$			*ns*	$p < .01$
	Vocabulary				
Lower academic track	40.79 (8.98)	44.86 (10.43)	49.29 (9.44)	4.07 (8.34)	8.51 (8.03)
Middle academic track	47.06 (7.86)	51.92 (10.03)	54.86 (9.76)	4.86 (8.51)	7.79 (9.02)
Upper academic track	54.34 (7.70)	59.37 (7.33)	61.59 (6.99)	5.04 (7.40)	7.25 (8.01)
Full sample	49.88 (9.72)	54.68 (10.48)	57.51 (9.63)	4.80 (7.87)	7.63 (8.25)
Test of significance [a]	$p < .01^{b}$			*ns*	*ns*

Note. The estimates refer to the subsample of all secondary school students for whom data from the elementary school years were available. Furthermore, students were still actively participating in the BiKS study in Grade 7, did not change their type of school, and had not repeated a class during the time period under investigation (active sample). Sample size was n = 136 students in lower academic track schools, n = 150 in middle academic track schools, and n = 372 students in upper academic track schools.

[a]It was tested whether estimates were equal between students attending lower, middle and upper academic track schools.

[b]Mplus Type is General was used as Grade 5 reading comprehension/vocabulary was treated as manifest.

Due to unequal sample sizes of the students attending different school tracks in the current sample and the special interest in the effect of attending upper academic track schools, in which the curriculum has a strong focus on preparing students for university entrance, in comparison to lower and middle academic track schools, which both mainly focus on preparing students for vocational training, students attending the lower and middle academic track schools were combined into one comparison group. Therefore, the analyses that were conducted by using Propensity-Score-Matching (PSM) focused on the estimation of the effect of attending an upper academic track school in comparison to attending lower or middle academic track schools between Grade 5 and Grade 7 on the development of reading comprehension and vocabulary. A broad set of covariates was used in order to adequately control for the treatment assignment. Radius matching with caliper was used as the matching procedure.

The distribution of the estimated propensity scores for students attending the lower and middle academic track schools (the controls) and students attending upper academic track schools is depicted in Figure 4 (without taking mathematics and German grades into account) and Figure 5 (after additionally taking mathematics and German grades into account). A graphical inspection of Figure 4 indicates that the distribution of propensity scores for students attending the lower and middle academic track schools was highly positive or right-skewed, whereas the distribution of the propensity scores of the upper academic track students was highly negative or left-skewed. Nevertheless, the figure also indicates that in between the two peaks, a relatively large region of overlap between the two distributions was present. Therefore, we expected a satisfactory number of comparable students for the matching procedure in the two groups and a good extrapolation with regard to the interpretation of the estimated results. By contrast, regarding the distribution of the propensity scores in Figure 5, when additionally considering mathematics and German grades of the students in Grade 4, it becomes obvious that the region of overlap decreased substantially. This can be seen by the lower number of students of the two groups who fell into the middle region or region of overlap when comparing Figure 5 with Figure 4. This effect is mainly attributable to the fact that in the state of Bavaria in particular, school choice is almost directly linked to the students' grades in Grade 4. Therefore, estimations of the effect of attending an upper academic track school in comparison to lower and middle academic track schools that take students' mathematics and German grades into account might be less affected by systematic biases due to unconsidered covariates but at the price of a lower extrapolation of the results to a larger population of students.

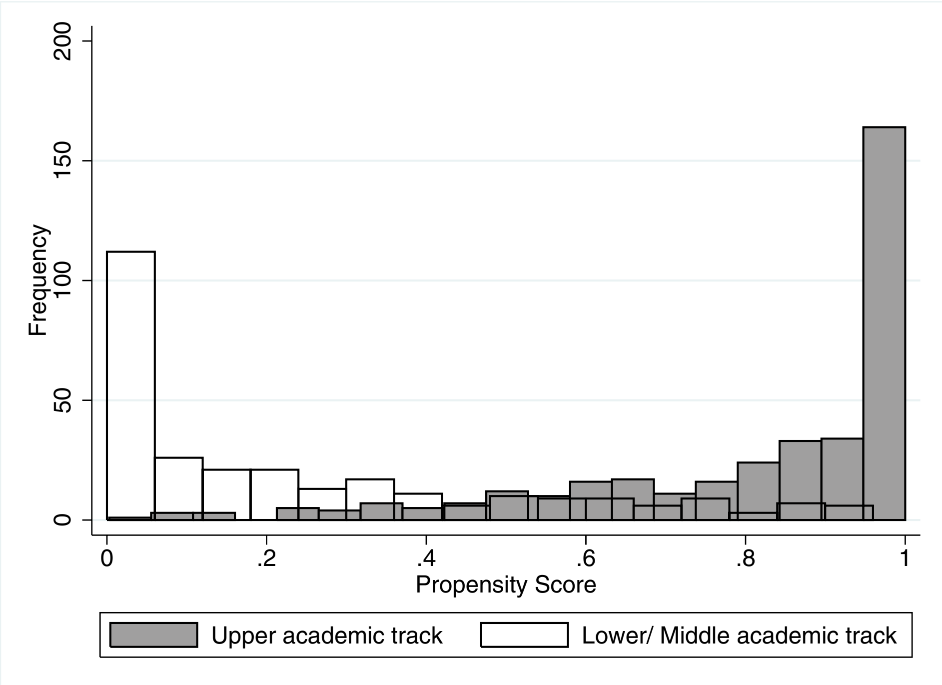

Figure 4. Distribution of propensity scores by school track without taking grades into account. Before matching, active sample: *M*(Upper academic track students) = 0.817; *M*(Lower/Middle academic track students) = 0.239; Standardized Difference = 234.1%; After radius matching: *M*(Upper academic track students) = 0.709; *M*(Lower/Middle academic track students) = 0.708; Standardized Difference = 0.1%.

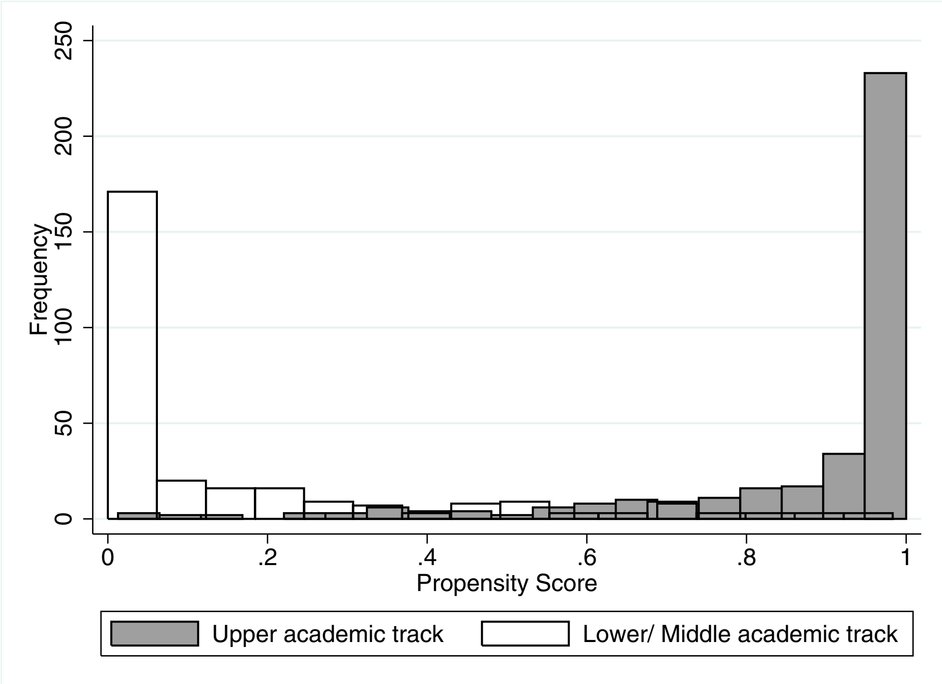

Figure 5. Distribution of propensity scores by school track after taking grades into account. Before matching, active sample: *M*(Upper academic track students) = 0.882; *M*(Lower/Middle academic track students) = 0.154; Standardized Difference = 326.9%; After radius matching: *M*(Upper academic track students) = 0.757; *M*(Lower/Middle academic track students) = 0.757; Standardized Difference = 0.0%.

In the next step, the balance with regard to the covariates between the two groups before and after the matching procedure was checked (Table 4). In the unmatched full sample, the estimates clearly indicated marked differences in the characteristics of the students who entered the upper academic track schools in comparison to the students who entered the lower and middle academic track schools (first column). Students attending upper academic track schools on average came more often from the federal state of Hesse, were younger, had better educated parents, came from families possessing more economic and cultural capital, read more in their leisure time, had a higher reading self-concept, and performed better on a wide range of achievement tests (vocabulary, mathematics, spelling, general cognitive abilities, and reading comprehension) in Grade 4 of elementary school. Finally, large differences in the German and mathematics grades in Grade 4 were present. After the first matching procedure, differences between the two groups of students were reduced substantially on most variables. However, some significant differences, especially on the categorical dummy-coded variables and the immigration background of the students remained, reflecting problems due to the small sample size in combination with large differences on several characteristics between students attending different school tracks. Furthermore, substantial differences in the German, mathematics, and science grades in Grade 4 remained, as these three variables were not included as covariates in the matching procedure.

Table 4. Covariate Imbalance in Unmatched and Matched Samples

Factor	Before matching[1]	Matched, without grades[1]	Matched, grades included[1,6]
State (1 = Bavaria)[2]	-48.3**	-14.9	-27.5*
Sex (1 = male)[2]	-13.0	-2.9	2.2
Age	-41.8**	7.2	-0.2
Education Parents [2,3]	117.0**	6.7	-11.8
Immigration (1 = immigration background) [2]	10.7	20.6*	22.4*
HISEI	104.1**	-7.9	-16.8
Cultural capital category 3[2]	-48.4**	9.7	0.2
Cultural capital category 4[2]	-28.3**	-11.1	21.7
Cultural capital category 5[2]	-16.8*	9.3	-4.6
Cultural capital category 6[2]	28.4**	-23.3*	-22.4
Cultural capital category 7[2]	51.8**	16.6	11.4
Reading behavior category 2[2,4]	-17.5*	4.1	-16.8
Reading behavior category 3[2,4]	-25.8**	-0.4	4.3
Reading behavior category 4[2,4]	-25.6**	8.0	21.6*
Reading self-concept category 2[2]	-24.2**	-16.8	-10.3
Reading self-concept category 3[2]	-35.8**	22.2*	24.7
Reading self-concept category 4[2]	51.8**	-15.0	-19.4
Vocabulary	101.1**	-5.3	-15.9
Mathematics competence	87.4**	12.7	22.4
Spelling	114.2**	-15.6*	-8.6
General cognitive abilities	63.9**	1.3	-4.1
Reading comprehension	100.4**	-8.3	8.3
Mathematics grades[5]	-134.9**	-73.8**	6.5
German grades[5]	-193.8**	-96.3**	1.8
Science grades[5]	-137.0**	-56.0**	9.1
Mean value[7]	64.9	18.6	12.6

Note. Standardized differences in percent (%). Formula from Rosenbaum and Rubin (1985).
[1] In general, a positive algebraic sign indicates a higher mean value in the treatment group (= upper academic school track); Results were computed using pstest implemented in psmatch2 (Leuven & Sianesi, 2003).
[2]The variable was dummy-coded.
[3]1 = parents reached university entrance qualification.
[4]Reading behavior was negatively keyed from category 1 = yes, every day to 4 = never or almost never;
[5]In Germany, grades are negatively keyed ranging from 1 = excellent to 6 = insufficient; the negative algebraic sign therefore indicates better (= lower) grades in the treatment group (= upper academic track).
[6]German and mathematics grades were included in the PSM; Science grades were not included as this led to severe imbalances on further covariates.
[7]All differences were treated as positive values.
* $p < .05$. ** $p < .01$.

The analyses of the outcome variables for the unmatched and matched samples, without taking school grades into account, are presented in Table 5. The results indicate that even after adjusting for a broad set of covariates, significant differences remained in reading comprehension and vocabulary between students attending upper academic track schools and students attending lower or middle academic track schools. For reading comprehension, the estimated effect of attending 3 years of an upper academic track school was about d = 0.33 in the matched sample. With regard to the development of vocabulary, an effect of d = 0.34 was estimated. The effect just missed the 5% significance level, but the sample size had been substantially reduced due to the matching. However, when considering German and mathematics grades in Grade 4 as additional covariates, the results changed (Table 6). Whereas in the first matching, substantial differences in the matched groups in German, mathematics, and science grades were still present, the second analysis also achieved a satisfactory balance on these three covariates (Table 4). However, the balance on most other covariates was less satisfactory. Furthermore, as already mentioned, the number of students within the region of common support and to whom the analyses referred decreased substantially after the inclusion of the German and mathematics grades (from n = 351 to n = 170; cf. Figures 4 and 5). With regard to the outcome – the development of reading comprehension – the estimated average treatment effect for the treatment group was d = 0.48. For the second outcome – vocabulary – the results of the radius matching did not indicate a significant difference between school types (d = 0.31).

Table 5. Reading Comprehension and Vocabulary in Grade 7 by School Track Before and After Matching

Outcome	Effect	*M* (upper academic track)	*M* (lower academic track)	*Diff.*	*SE*	*Diff*/ *SE*	*d*	Grade 4 *d*
Reading comprehension	Unmatched	61.343	46.873	14.470	1.064	13.595**	0.97	0.91
	Matched	58.052	53.129	4.923	2.177	2.261*	0.33	-0.08
Vocabulary	Unmatched	61.588	52.211	9.378	0.718	13.052**	0.97	0.92
	Matched	60.694	57.427	3.267	1.696	1.926	0.34	-0.05

Note. Grades were not included as covariates in the matching. Sample size was n = 658 students in the unmatched and n = 351 students in the Radius matched sample. *SD*(Reading comprehension, Grade 7) = 14.902; *SD*(Vocabulary, Grade 7) = 9.640.
* $p < .05$. ** $p < .01$.

Taken together, the results of the Propensity-Score-Matching analyses indicate a substantial positive effect of attending 3 years of an upper academic track school in comparison to lower and middle academic track schools. The estimated size of this effect varied from around d = 0.3 to d = 0.5 for reading comprehension as well as vocabulary. As mentioned, the selection process of attending the upper, middle, or lower academic tracks was, at least in the regions from where the present sample stemmed, strongly determined by the Grade 4 grades. However, grades are difficult to compare across different schools and classes, so taking these measures into account as covariates in the matching process might go along with imbalances on additional unobserved variables.

Table 6. Reading Comprehension and Vocabulary in Grade 7 by School Track Before and After Matching (incl. grades as covariates)

Outcome	Effect	*M* (upper academic track)	*M* (lower academic track)	*Diff.*	*SE*	*Diff/ SE*	*d*	Grade 4 *d*
Reading comprehension	Unmatched	61.343	46.873	14.470	1.064	13.595**	0.97	0.91
	Matched	59.850	52.633	7.218	3.400	2.123*	0.48	0.07
Vocabulary	Unmatched	61.588	52.211	9.378	0.718	13.052**	0.97	0.92
	Matched	60.855	57.899	2.956	2.749	1.075	0.31	-0.14

Note. Grades were considered as covariates in the matching procedure. Sample size was n = 658 students in the unmatched and n = 170 students in the Radius matched sample. *SD*(Reading comprehension, Grade 7) = 14.902; *SD*(Vocabulary, Grade 7) = 9.640.
* $p < .05$. ** $p < .01$.

Discussion

With regard to the first research question, the question of whether differences in the development of reading comprehension and vocabulary between different types of schools or school tracks could be found, the analyses showed a widening gap between students attending upper, middle, and lower academic track schools in reading comprehension between Grade 5 and Grade 7. Furthermore, the effect of increasing differences in reading comprehension was demonstrated independently of the treatment of student dropout in the analytic model. Therefore, the developmental pattern of reading comprehension in the first years of elementary school fits well with the notion of a fan-spread effect and converges well with results that have been reported in the domain of mathematics (Becker, et al., 2006; Köller & Baumert, 2001; Schmidt, 2009) but contrast with findings often reported in reading (Gröhlich, et al., 2009; Lehmann, et al., 1998; Retelsdorf & Möller, 2008).

In the domain of vocabulary, the findings did not support the assumption of a widening gap between different types of schools. Furthermore, results differed slightly by the different treatment of student dropout: Analyses that ignored student dropout by imputing all missing values indicated a small, although significant catch-up effect for students attending lower academic track schools, whereas analyses that excluded all students who were no longer participating in the last wave of measurement found stable differences in vocabulary between the three different school tracks. When taking a closer look at the differences between the estimated values of these two analyses, we see that the subsample of the "survivors" (students who still active participate in the study in Grade 7) in general scored higher on measures of reading comprehension and vocabulary than the full sample, indicating that lower competence is linked to an increased probability of student dropout. Furthermore, this tendency was moderated by the school track: Whereas student dropout was almost not or only slightly positively linked to achievement measures in lower academic track schools, student dropout was negatively linked to achievement differences in middle and upper academic track schools. These differences might be attributable to characteristics of the school system: Whereas in upper academic track schools, students can change only to a less demanding school type, students in lower academic track schools can additionally change to more demanding school types. Taken together, the vocabulary gap between students staying in the different school tracks (and therefore still active participating in the BiKS-study) seemed to remain stable. Slightly higher vocabulary trends however were estimated for students leaving the lower track (and therefore in most cases dropping-out of the study), indicating the need for further research dedicated to the analyses of developmental trends for students changing school track.

But why did differences in vocabulary remain more or less stable, whereas differences in reading comprehension between school tracks tend to increase with time? There are at least two explanations for this result. According to a technical explanation, differences in the development of reading comprehension and vocabulary might be an artifact of different test characteristics. Tests might differ in their sensitivity to detect changes in the latent trait. The second explanation, an educational explanation, assumes that differences in the learning mechanisms are responsible for these developmental differences. Whereas vocabulary knowledge may be mostly acquired subconsciously by processes of incidental learning (Krashen, 1989), the fostering of reading comprehension may still be explicitly due to instruction in school. As a consequence, measures of reading comprehension should be more sensitive to between-school differences due to institutional differences in the content and quality of instruction. Nevertheless, this explanation is only partially supported by the findings of the second set of analyses, which will be discussed next.

What is the Effect of Attending an Upper Academic Track School on Learning?

Tracing interindividual differences in learning between different school tracks does not instantaneously mean that these differences are the product of different learning environments. Rather, differences in learning rates between different types of schools or school tracks might arise from the interplay of institutional characteristics with differences in the composition of the students and the individual traits and abilities of the students that already exist prior to the attendance of secondary school (Ditton & Krüsken, 2006; Pfost, Karing, et al., 2010; Schneider & Stefanek, 2004). Disentangling these different sources is of special scientific interest, but creating experimental conditions in which students can be randomly assigned to different school tracks is not feasible. The BiKS study, however, provides analytic possibilities for addressing this question because data on the students who attend different secondary school tracks are available, and these data have already been measured in elementary school (prior to the treatment exposure). To make use of this favorable circumstance in the current study, Propensity-Score-Matching as a tool for analyzing treatment effects in nonequivalent treatment groups was applied. In order to control for selectivity into the different secondary schools, a broad number of factors, including achievement measures from Grade 4, which might influence students' school choice or the outcome, were taken into account as covariates. Students' school grades in German and mathematics in the middle of Grade 4 were considered in an additional analysis, but their use went along with the loss of a broad number of matches. Furthermore, school grades are often not directly comparable beyond classes, schools, and regions because teachers are inveigled into using different reference scales (Maaz, et al., 2008; Trautwein, et al., 2008; Treutlein & Schöler, 2009). Science grades were not included as an additional covariate. A model that included the grades of all three main subjects (German, mathematics, and science) led to a strong imbalance on most covariates and was therefore not considered. Although not included as a covariate, differences in science grades between the different school tracks were nevertheless substantially reduced by the applied Propensity-Score-Matching.

The results of the matching analyses that had not taken school grades into account as a covariate indicated a positive effect of attending an upper academic track school on the development of reading comprehension and vocabulary (the effect for vocabulary slightly missed the 5% significance level but was still substantial in terms of effect size). Regarding the magnitude of the effect on reading comprehension and vocabulary across a 3-year period, from the end of Grade 4 to Grade 7, students in upper academic track schools gained about one third of a standard deviation more than we expected that they would have learned when attending lower and middle academic track schools (the estimated counterfactual outcome). When taking grades in mathematics and German into account as further covariates, this positive significant effect of attending an upper academic track school on learning did not change substantially for reading comprehension. For vocabulary there was as strong increase in the standard error, so the effect was far away from reaching statistical significance although just marginally changing in terms of effect size. This means that although the null hypothesis of equal development between the matched pairs who attended different school tracks could not be rejected, differences in the sample that were not negligible in size remained. Comparing this cumulative 3-year effect to an empirical benchmark indicated that the emerging difference between the end of Grade 4 and Grade 7 in our sample was comparable to the normative change we would expect in the domain of reading from at least a half year of schooling (Bloom, Hill, Black, & Lipsey, 2008; Hill, et al., 2008).

So, taken together, what do the results of the matching analysis tell us? First, results need to be interpreted against the background of the assumptions underlying the analysis. As long as unobserved or unconsidered covariates that influence the treatment assignment as well as the treatment outcome and that have not been blocked by conditioning on the considered covariates are present, results may be systematically biased. In the current study, we tried to map the process of selecting a certain school track by taking a set of prominent covariates into account. Nevertheless, it should be acknowledged that the real process of selecting a certain type of school might be much more complex than assumed in the present analyses. And second, the role of school grades as a confounding factor between school choice and competence development beyond objective achievement measures, measures of the economic, ethnic, and familial background of the students, as well as further individual characteristics of students need further investigation. Thereby, we should ask about the appropriateness of using measures such as school grades that differ in meaning between subjects due to differential context conditions.

Limitations

Analyzing the development of reading literacy in the different school tracks is a sensitive topic that needs to be treated cautiously. Analyses are sensitive to the subjects who are considered. Student dropout in longitudinal studies may occur for meaningful reasons such as a change in school type, moving to another city, the repetition of a grade, and so on (van de Grift, 2009). Therefore, in the analysis of fan-spread effects the treatment of missing values may become a central theme that has to be taken into account. In our first model, reading comprehension and vocabulary development were analyzed under the assumption that no change in the type of school occurred during the period under investigation. All missing values regardless of participation status were estimated by multiple imputation. However, we should keep in mind that student dropout was quite substantial, as only 1,358 out of 1,801 (75.4%) secondary school students participated in Grade 7 (additionally, for 120 participating students, competence measures were missing in Grade 7). Imputation of such large amounts of missing data might be critical and might explain by itself the differences found in estimated growth when compared to the students who were still actively participating. Consequently, the same analysis was run by considering only the students who were still present in Grade 7 – the active sample (N = 1,358). Nevertheless, both approaches neglected the dynamic character of the students who remained but also changed schools. Additionally, the present analyses were limited to students whose parents decided to actively participate in the BiKS study (active informed consent). Within the

BiKS study, students with an immigration background as well as students with higher (i.e., worse) grades were underrepresented in the sense that these students (i.e., their parents) more frequently actively or passively refused to participate in the study (cf. Pfost, 2011). Therefore, the current sample was not fully representative of all students from the participating schools or of all students in the federal states of Bavaria and Hesse.

Another limitation of this study concerns the measurement and scaling of reading comprehension. In the current study, reading comprehension was measured by using different items at different waves of measurement in combination with items that were presented to the students a second time (common item design with nonequivalent groups/ anchor-item test design: Holland, et al., 2007; Kolen & Brennan, 2004), and students' reading comprehension was estimated on a common metric by using a logit-link function within an IRT framework. However, equating across grade levels (vertical scaling) in particular may produce different results depending on the equating methodology used in combination with substantial equating error, particularly when assumptions of the measurement model are not met (Wu, 2010). A new presentation of identical test material, as practiced in the domain of vocabulary, does not necessarily solve scaling problems and may create additional problems such as memory effects. Thus, in summary, as long as we do not have natural metrics, research findings may be substantially biased by scaling artifacts (Embretson, 2006).

Finally, it should be noted that Propensity-Score-Matching is only a weak alternative for the analysis of treatment effects in comparison to randomized experiments. PSM can adjust only for observed confounding covariates, whereas randomization tends to balance the distribution of all covariates, observed and unobserved (Rubin, 1997). Therefore, the estimated effects of attending an upper academic track school in comparison to lower or middle academic track schools can be interpreted only against the background of covariates that were taken into account and for which balance between the matched samples could be achieved. Furthermore, the estimated results can only be interpreted as a narrower treatment effect, the common-support treatment effect for the treated (Morgan & Winship, 2007). This means that, even if the assumption of conditional ignorability was true in the present case, the estimated effect refers only to those students who typically get the treatment, which means students who would typically choose an upper academic track school and for whom valid counterparts in the control condition could be found. Or, in simpler terms, the estimates refer primarily to those students for whom the choice of type of school after Grade 4 was not perfectly determined by their performance, ethnic or social background, and so forth. Further discussion and assumptions concerning the causal interpretability of estimated results in observational studies are presented in Morgan and Winship (2007), Rubin (1986, 2004), Shadish (2010), and West and Thoemmes (2010). To conclude, although estimations of the effect of attending different school tracks on the development of reading comprehension and vocabulary tried to take into account a broad set of potential confounding variables that have been observed in the BiKS study in combination with up-to-date analytical methods, all estimated results should be interpreted with great caution and after reflecting upon the underlying assumptions.

Implications for Future Research

Tracing the development of cognitive competencies in different types of schools or school tracks with observational studies is a very sensitive topic. Therefore, future research should devote more resources toward further improving studies with regard to the measures used, the scaling techniques applied, and the sample selected for observation. On the other hand, estimating the effect of attending different school tracks on the development of cognitive competencies does not tell us anything about the mechanisms that mediate these effects. Therefore, beyond asking how successful schools are in promoting the cognitive development of students, we further need to ask why these differences occur. And finally, we may be interested in the question of the fit between the type of school and student characteristics. Effects of attending different school tracks may vary for different subpopulations of students, a topic that needs further attention in future research.

References

Aarnoutse, C., van Leeuwe, J., Voeten, M., & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing, 14*, 61-89. doi:10.1023/A:1008128417862

Ariga, K., & Brunello, G. (2007). *Does secondary school tracking affect performance? Evidence from IALS*. Bonn: Forschungsinstitut zur Zukunft der Arbeit.

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of matthew effects in reading: Results from a dutch longitudinal study. *Developmental Psychology, 34*, 1373-1399. doi:10.1037/0012-1649.34.6.1373

Baumert, J. (2006). Was wissen wir über die Entwicklung von Schulleistungen? *Pädagogik, 58*, 40-46.

Baumert, J., Becker, M., Neumann, M., & Nikolova, R. (2009). Frühübergang in ein grundständiges Gymnasium – Übergang in ein privilegiertes Entwicklungsmilieu? Ein Vergleich von Regressionsanalyse und Propensity Score Matching. *Zeitschrift für Erziehungswissenschaft, 12*, 189-215. doi:10.1007/s11618-009-0072-4

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, J. M., et al. (Eds.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.

Baumert, J., & Köller, O. (2005). Sozialer Hintergrund, Bildungsbeteiligung und Bildungsverläufe im differenzierten Sekundarschulsystem. In V. Frederking, H. Heller & A. Scheunpflug (Eds.), *Nach PISA. Konsequenzen für Schule und Lehrerbildung nach zwei Studien* (pp. 9-21). Wiesbaden: VS Verlag für Sozialwissenschaften.

Baumert, J., Köller, O., & Schnabel, K. (1999). Schulformen als differentielle Entwicklungsmilieus – eine ungehörige Fragestellung? Erwiderung auf die Expertise "Zur Messung sozialer Motivation in der BIJU-Studie" von Georg Lind. In G. E. u. Wissenschaft (Ed.), *Messung sozialer Motivation: Eine Kontroverse* (pp. 28-69). Frankfurt am Main: Bildungs- und Förderwerk der Gewerkschaft Erziehung und Wissenschaft im DGB e. V.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*, 133-180. doi:10.3102/0002831209345157

Baumert, J., & Schümer, G. (2001). Schulformen als selektionsbedingte Lernmilieus. 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. 454-467). Opladen: Leske + Budrich.

Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In J. Baumert, P. Stanat & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungssystem* (pp. 95-188). Wiesbaden: VS Verlag für Sozialwissenschaften.

Becker, M., Lüdtke, O., Trautwein, U., & Baumert, J. (2006). Leistungszuwachs in Mathematik. Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem? *Zeitschrift für Pädagogische Psychologie, 20*, 233-242. doi:10.1024/1010-0652.20.4.233

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*, 289-328. doi:10.1080/19345740802400072

Burns, M. S., & Kidd, J. K. (2010). Learning to read. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education* (3 ed., pp. 394-400). Oxford, United Kingdom: Elsevier.

Chall, J. S. (1983). *Stages of reading development.* New York: McGraw-Hill.

Cortina, K. S., Baumert, J., Leschinsky, A., Mayer, K. U., & Trommer, L. (Eds.). (2008). *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick*. Reinbek bei Hamburg: Rowohlt.

Cortina, K. S., & Trommer, L. (2005). Bildungswege und Bildungsbiographien in der Sekundarstufe 1. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Eds.), *Das Bildungswesen in der Bundesrepublik Deutschland* (pp. 342-391). Reinbek bei Hamburg: Rowohlt.

Courser, M. W., Shamblen, S. R., Lavrakas, P. J., Collins, D., & Ditterline, P. (2009). The impact of active consent procedures on nonresponse and nonresponse error in youth survey data. Evidence from a new experiment. *Evaluation Review, 33*, 370-395. doi:10.1177/0193841X09337228

Dang, H.-A., & Rogers, F. H. (2008). *How to interpret the growing phenomen of private tutoring: Human capital deepening, inequality increasing, or waste of resources?* Policy Research Working Paper 4530. The World Bank. Retrieved from http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2008/02/25/000158349_20080225153509/Rendered/PDF/wps4530.pdf

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics, 84*, 151-161.

Ditton, H., & Krüsken, J. (2006). Der Übergang von der Grundschule in die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft, 9*, 348-372. doi:10.1007/s11618-006-0055-7

Ditton, H., Krüsken, J., & Schauenberg, M. (2005). Bildungsungleichheit – der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft, 8*, 285-304. doi:10.1007/s11618-005-0138-x

Dreeben, R., & Barr, R. (1988). Classroom Composition and the design of instruction. *Sociology of Education, 61*, 129-142.

Dumay, X., & Dupriez, V. (2007). *Does the school composition effect matter? Some methodological and conceptional considerations*. Les Cahiers de Recherche en Éducation et Formation, (60). Université Catholique de Louvain, Louvain-la-Neuve.

Embretson, S. E. (2006). The continued search for nonarbitrary metrics in psychology. *American Psychologist, 61*, 50-55. doi:10.1037/0003-066X.61.1.50

Esbensen, F. A., Deschenes, E. P., Vogel, R. E., West, J., Arboit, K., & Harris, L. (1996). Active parental consent in school-based research. An examination of ethical and methodological issues. *Evaluation Review, 20*, 737-753. doi:10.1177/0193841X9602000605

Esbensen, F. A., Hughes Miller, M., Taylor, T. J., He, N., & Freng, A. (1999). Differential attrition rates and active parental consent. *Evaluation Review, 23*, 316-335. doi:10.1177/0193841X9902300304

Faust, G. (2005). Übergänge in den Sekundarbereich. In W. Einsiedler, M. Götz, H. Hacker, J. Kahlert, R. W. Keck & U. Sandfuchs (Eds.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (pp. 291-296). Bad Heilbrunn: Klinkhardt.

Faust, G. (2006). Zum Stand der Einschulung und der neuen Schuleingangsstufe in Deutschland. *Zeitschrift für Erziehungswissenschaft, 9*, 328-347. doi:10.1007/s11618-006-0054-8

Frey, D., Heinze, A., Mildner, D., Hochweber, J., & Asseburg, R. (2010). Mathematische Kompetenz von PISA 2003 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt* (pp. 153-176). Münster: Waxmann.

Gamoran, A., & Berends, M. (1987). The effects of stratification in secondary schools: synthesis of survey and ethnographic research. *Review of Educational Research, 57*, 415-435. doi:10.3102/00346543057004415

Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: compensation, reinforcement, or neutrality? *American Journal of Sociology, 94*, 1146-1183.

Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research, 21*, 1-56.

Geiser, C. (2010). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Gölitz, D., Roick, T., & Hasselhorn, M. (2005). Deutsche Mathematiktests für dritte und vierte Klassen (DEMAT 3+ und DEMAT 4). In M. Hasselhorn, H. Marx & W. Schneider (Eds.), *Diagnostik von Mathematikleistungen* (Vol. Band 4, pp. 167-198). Göttingen: Hogrefe.

Gröhlich, C., Bonsen, M., & Bos, W. (2009). Von KESS 4 zu KESS 7: Lernentwicklung in der Beobachtungsstufe. In W. Bos, M. Bonsen & C. Gröhlich (Eds.), *KESS 7. Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen zu Beginn der Jahrgangsstufe 7* (pp. 91-122). Münster: Waxmann.

Grund, M., Haug, G., & Naumann, C. L. (2003). *DRT 4. Diagnostischer Rechtschreibtest für 4. Klassen*. Göttingen: Beltz.

Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School effectiveness and school improvement, 15*, 177-199. doi:10.1076/sesi.15.2.177.30432

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

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*, 172-177. doi:10.1111/j.1750-8606.2008.00061.x

Holland, P. W., Dorans, N. J., & Peterson, N. S. (2007). Equating test scores. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics 26. Psychometrics* (pp. 169-203). Amsterdam: Elsevier.

Kempe, C., Eriksson-Gustavsson, A.-L., & Samuelsson, S. (2011). Are there any Matthew effects in literacy and cognitive development. *Scandinavian Journal of Educational Research, 55*, 181-196. doi:10.1080/00313831.2011.554699

Klicpera, C., Schabmann, A., & Gasteiger-Klicpera, B. (1993). Lesen- und Schreibenlernen während der Pflichtschulzeit: Eine Längsschnittuntersuchung über die Häufigkeit und Stabilität von Lese- und Rechtschreibschwierigkeiten in einem Wiener Schulbezirk. *Zeitschrift für Kinder- und Jugendpsychiatrie, 21*, 214-225.

Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., et al. (Eds.). (2010). PISA 2009. Bilanz nach einem Jahrzehnt. Münster: Waxmann.

Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking. Methods and Practices*. New York, NY: Springer.

Köller, O., & Baumert, J. (2001). Leistungsgruppierungen in der Sekundarstufe 1. Ihre Konsequenzen für die Mathematikleistung und das mathematische Selbstkonzept der Begabung. *Zeitschrift für Pädagogische Psychologie, 15*, 99-110. doi:10.1024//1010-0652.15.2.99

Köller, O., & Baumert, J. (2008). Entwicklung schulischer Leistungen. In R. Oerter & L. Montada (Eds.), *Entwicklungspsychologie* (pp. 735-768). Weinheim: Beltz.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis. *The Modern Language Journal, 73*, 440-464.

Kunter, M., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W., et al. (2005). Der Mathematikunterricht der PISA-Schülerinnen und Schüler. *Zeitschrift für Erziehungswissenschaft, 8*, 502-520. doi:10.1007/s11618-005-0156-8

Lehmann, R. H. (2006). Zur Bedeutung der kognitiven Heterogenität von Schulklassen für den Lernstand am Ende der Klassenstufe 4. In A. Schründer-Lenzen (Ed.), *Risikofaktoren kindlicher Entwicklung. Migration, Leistungsangst und Schulübergang*. (pp. 109-121). Wiesbaden: VS Verlag für Sozialwissenschaften.

Lehmann, R. H., & Lenkeit, J. (2008). *ELEMENT. Erhebung zum Lese- und Mathematikverständnis. Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin*. Berlin: Senatsverwaltung für Bildung, Jugend und Sport.

Lehmann, R. H., Peek, R., Gänsfuß, R., & Hußfeldt, V. (1998). *LAU 7. Aspekte der Lernausgangslage und der Lernentwicklung - Klassenstufe 7. Ergebnisse einer längsschnittlichen Untersuchung in Hamburg*. Hamburg: Behörde für Bildung und Sport.

Lenhard, W., & Schneider, W. (2005). *ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässler*. (1. Auflage ed.). Göttingen: Hogrefe.

Leschinsky, A. (2008a). Die Hauptschule - von der Be- zur Enthauptung. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Eds.), *Das Bildungswesen in der Bundesrepublik Deutschland* (pp. 377-406). Reinbek bei Hamburg: Rowohlt.

Leschinsky, A. (2008b). Die Realschule - ein zweischneidiger Erfolg. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Eds.), *Das Bildungswesen in der Bundesrepublik Deutschland* (pp. 407-436). Reinbek bei Hamburg: Rowohlt.

LeTendre, G. K., Hofer, B. K., & Shimizu, H. (2003). What is tracking? Cultural expectations in the United States, Germany, and Japan. *American Educational Research Journal, 40*, 43-89. doi:10.3102/00028312040001043

Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing (Version 4.0.4). Retrieved from http://ideas.repec.org/c/boc/bocode/s432001.html

Ma, X., & Klinger, D. A. (2000). Hierarchical linear modelling of student and school effects on academic achievement. *Canadian Journal of Education, 25*, 41-55.

Maaz, K., Hausen, C., McElvany, N., & Baumert, J. (2006). Stichwort: Übergänge im Bildungssystem. Theoretische Konzepte und ihre Anwendung in der empirischen Forschung beim Übergang in die Sekundarstufe. *Zeitschrift für Erziehungswissenschaft, 9*, 299-327. doi:10.1007/s11618-006-0053-9

Maaz, K., Neumann, M., Trautwein, U., Wendt, W., Lehmann, R. H., & Baumert, J. (2008). Der Übergang von der Grundschule in die weiterführende Schule: die Rolle von Schüler- und Klassenmerkmalen beim Einschätzen der individuellen Lernkompetenz durch die Lehrkräfte. *Schweizerische Zeitschrift für Bildungswissenschaften, 30*, 519-548.

McElvany, N., Kortenbruck, M., & Becker, M. (2008). Lesekompetenz und Lesemotivation. Entwicklung und Mediation des Zusammenhangs durch Leseverhalten. *Zeitschrift für Pädagogische Psychologie, 22*, 207-219. doi:10.1024/1010-0652.22.34.207

Morgan, P. L., & Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Exceptional children, 73*, 165-183.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference. Methods and principles for social research*. Cambridge, NY: Unversity Press.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide*. (6 ed.). Los Angeles, CA: Muthén & Muthén.

Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt* (pp. 23-71). Münster: Waxmann.

Opdenakker, M.-C., van Damme, J., de Fraine, B., van Landeghem, G., & Onghena, P. (2002). The effect of schools and classes on mathematics achievement. *School effectiveness and school improvement, 13*, 399-427. doi:10.1076/sesi.13.4.399.10283

Parrila, R., Aunola, K., Leskinen, E., Nurmi, J.-E., & Kirby, J. (2005). Development of individual differences in reading: results from longitudinal studies in English and Finnish. *Journal of Educational Psychology, 97*, 299-319. doi:10.1037/0022-0663.97.3.299

Pfost, M. (2011). *Klassenkompositionseffekte in der Sekundarstufe und Prüfung vermittelnder Mechanismen*. Unveröffentlichte Dissertationsschrift, Otto-Friedrich-Universität Bamberg.

Pfost, M., Dörfler, T., & Artelt, C. (2010). Der Zusammenhang zwischen außerschulischem Lesen und Lesekompetenz. Ergebnisse einer Längsschnittstudie am Übergang von der Grund- in die weiterführende Schule. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42,* 167-176. doi:10.1026/0049-8637/a000017

Pfost, M., Dörfler, T., & Artelt, C. (2012). Reading competence development of poor readers in a German elementary school sample. An empirical examination of the Matthew-effect model. *Journal of Research in Reading, 35,* 411-426. doi:10.1111/j.1467-9817.2010.01478.x

Pfost, M., Karing, C., Lorenz, C., & Artelt, C. (2010). Schereneffekte im ein- und mehrgliedrigen Schulsystem. Differenzielle Entwicklung sprachlicher Kompetenzen am Übergang von der Grund- in die weiterführende Schule? *Zeitschrift für Pädagogische Psychologie, 24,* 259-273. doi:10.1024/1010-0652/a000025

Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., et al. (Eds.). (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie.* Münster: Waxmann.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R. H., Leutner, D., Neubrand, M., et al. (Eds.). (2005). *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* Münster: Waxmann.

Renkl, A. (1996). Vorwissen und Schulleistung. In J. Möller & O. Köller (Eds.), *Emotionen, Kognitionen und Schulleistung* (pp. 175-190). Weinheim: Psychologie Verlags Union.

Retelsdorf, J., & Möller, J. (2008). Entwicklungen von Lesekompetenz und Lesemotivation. Schereneffekte in der Sekundarstufe? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 40,* 179-188.

Rönnebeck, S., Schöps, K., Prenzel, M., Mildner, D., & Hochweber, J. (2010). Naturwissenschaftliche Kompetenz von PISA 2006 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt* (pp. 177-198). Münster: Waxmann.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistican, 39*, 33-38. doi:10.2307/2683903

Rubin, D. B. (1986). Which ifs have causal answers. *Journal of the American Statistical Association, 81*, 961-962.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine, 127*, 757-763.

Rubin, D. B. (2004). Teaching statistical inference for causal effects in experiments and observatinal studies. *Journal of Educational and Behavioral Statistics, 29*, 343-367. doi:10.3102/10769986029003343

Schmidt, S., Schmitt, M., & Smidt, W. (2009). *Die BiKS-Studie. Methodenbericht zur zweiten Projektphase*. Bamberg: Otto-Friedrich-Universität.

Schmidt, W. H. (2009). *Exploring the relationship between content coverage and achievement: unpacking the meaning of tracking in eight grade mathematics*. Michigan: The Education Policy Center at Michigan State University.

Schneider, T. (2004). Nachhilfe als Strategie zur Verwirklichung von Bildungszielen. Eine empirische Untersuchung mit Daten des Sozio-oekonomischen Panels (SOEP) *Discussion papers 447*. Berlin: DIW.

Schneider, W., & Stefanek, J. (2004). Entwicklungsveränderungen allgemeiner kognitiver Fähigkeiten und schulbezogener Fertigkeiten im Kindes- und Jugendalter. Evidenz für einen Schereneffekt? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 36*, 147-159. doi:10.1026/0049-8637.36.3.147

Shadish, W. R. (2010). Campbell and Rubin: a primer and comparison of their approaches to causal inference in field settings. *Psychological Methods, 15*, 3-17. doi:10.1037/a0015916

Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research, 60*, 471-499. doi:10.3102/00346543060003471

Spear-Swerling, L., Brucker, P. O., & Alfano, M. P. (2010). Relationships between sixth-graders' reading comprehension and two different measures of print exposure. *Reading and Writing, 23*, 73-96. doi:10.1007/s11145-008-9152-8

Stanat, P. (2006). Schulleistungen von Jugendlichen mit Migrationshintergrund: Die Rolle der Zusammensetzung der Schülerschaft. In J. Baumert, P. Stanat & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000* (pp. 189-219). Wiesbaden: Verlag für Sozialwissenschaften.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407. doi:10.1598/RRQ.21.4.1

Stanovich, K. E. (2000). *Progress in understanding reading. Scientific foundations and new frontiers*. New York, NY: Guilford Press.

Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true individual change: true change as a latent variable. *Methods of Psychological Research Online, 2,* 21-33.

Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. *Statistical Science, 25,* 1-21. doi:10.1214/09-STS313

Trautwein, U., Lüdtke, O., Becker, M., Neumann, M., & Nagy, G. (2008). Die Sekundarstufe I im Spiegel der empirischen Bildungsforschung: Schulleistungsentwicklung, Kompetenzniveaus und die Aussagekraft von Schulnoten. In E. Schlemmer & H. Gerstberger (Eds.), *Ausbildungsfähigkeit im Spannungsfeld zwischen Wissenschaft, Politik und Praxis* (pp. 91-107). Wiesbaden: VS Verlag für Sozialwissenschaften.

Trautwein, U., & Neumann, M. (2008). Das Gymnasium. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Eds.), *Das Bildungswesen in der Bundesrepublik Deutschland* (pp. 467-501). Reinbek bei Hamburg: Rowohlt.

Treutlein, A., & Schöler, H. (2009). Zum Einfluss der schulischen Lernumwelt auf die Schulleistung. In J. Roos & H. Schöler (Eds.), *Entwicklung des Schriftspracherwerbs in der Grundschule* (pp. 109-143). Wiesbaden: VS Verlag für Sozialwissenschaften.

Unger, J. B., Gallaher, P. G., Palmer, P. H., Baezconde-Garbanati, L., Trinidad, D. R., Cen, S., et al. (2004). No news is bad news: Characteristics of adolescents who provide neither parental consent nor refusal for participation in school-based survey research. *Evaluation Review, 28*, 52-63. doi:10.1177/0193841X03254421

van Buuren, S., & Oudshoorn, C. G. M. (2000). *Multivariate Imputation by Chained Equations. MICE V1.0 User's manual*. Leiden: TNO Prevention and Health.

van de Grift, W. (2009). Reliability and validity in measuring the value added of schools. *School effectiveness and school improvement, 20*, 269-285. doi:10.1080/09243450902883946

van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*, 134-150. doi:10.1016/j.edurev.2010.02.001

Walter, O., & Stanat, P. (2008). Der Zusammenhang des Migrantenanteils in Schulen mit der Lesekompetenz: Differenzierte Analysen der erweiterten Migrantenstichprobe von PISA 2003. *Zeitschrift für Erziehungswissenschaft, 11*, 84-105. doi:10.1007/s11618-008-0005-7

Weiß, R. H. (1987). *Wortschatztest (WS) und Zahlenfolgentest (ZF). Ergänzungstests zum Grundintelligenztest CFT 20. Handanweisung*. Göttingen: Hogrefe.

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 - Revision - (CFT 20-R) mit Wortschatztest und Zahlenfolgentest - Revision (WS/ZF-R)*. Göttingen: Hogrefe.

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods, 15*, 18-37. doi:10.1037/a0015917

Wu, M. L. (2010). Measurement, sampling, and equating errors in large-scale assessments. *Educational Measurement: Issues and Practice, 29*, 15-27. doi:10.1111/j.1745-3992.2010.00190.x

Wu, M. L., Adams, R. J., Wilson, M., & Haldane, S. A. (2007). ACER ConQuest version 2.0: generalised item response modelling software. Camberwell: ACER Press.

Zimmer, R. W., & Toma, E. F. (2000). Peer effects in private and public schools across Countries. *Journal of Policy Analysis and Management, 19,* 75-92. doi:10.1002/(SICI)1520-6688(200024)19:1<75::AID-PAM5>3.0.CO;2-W

Chapter 9

9 Is Secondary School Teacher Judgment Accuracy Related to the Development of Students' Reading Literacy?[1]

Constance Karing, Maximilian Pfost, and Cordula Artelt

Summary

The present work focuses on the relation between teacher judgement accuracy and the development of students' reading literacy and whether this relation is moderated by instructional variables. Longitudinal data were obtained from a sample of 502 students and their 40 German language teachers in the context of the BiKS-8-14 longitudinal study (measurement points: at the ends of Grade 5 and

Author Note

Constance Karing,
Department of Research Synthesis, Intervention and Evaluation, University of Jena, Germany.

Maximilian Pfost,
Department of Educational Research, University of Bamberg, Germany.

Cordula Artelt,
Department of Educational Research, University of Bamberg, Germany.

This research was supported by grant AR 301/6-1 & AR 301/6-2 from the German Research Foundation (DFG).

Correspondence concerning this chapter should be addressed to Constance Karing, Department of Research Synthesis, Intervention and Evaluation, Humboldtstraße 26, 07743 Jena, Germany. E-mail: constance.karing@uni-jena.de

---

[1] The results reported in this book chapter rely on the article „Is there a relationship between lower secondary school teacher judgment accuracy and the development of students' reading and mathematical competence?" by Constance Karing, Maximilian Pfost, and Cordula Artelt, published first in the Journal for Educational Research Online (Karing, Pfost & Artelt, 2011).

Die in diesem Beitrag berichteten Ergebnisse beruhen im Wesentlichen auf dem Artikel „Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen?" von Constance Karing, Maximilian Pfost und Cordula Artelt, zuerst veröffentlicht im Journal for Educational Research Online (Karing, Pfost & Artelt, 2011).

Grade 6). Teacher judgement accuracy was measured by the task-specific hit rate and the rank-order component. German language teachers showed a moderate hit rate ($M = 0.66$). For the rank-order component, we found a mean correlation of $\bar{r} = .19$. Multilevel analyses revealed a significant positive relation between the task-specific hit rate and the development of students' reading literacy. Furthermore, this significant relation was moderated by instructional variables such as teachers' use of structural cues and the degree to which lessons were individualized. A high task-specific hit rate in combination with a high degree of individualization of lessons was significantly associated with an increased development in students' reading literacy. However, a high task-specific hit rate in combination with a low frequency of structural cue use during lessons was also significantly related to an increase in the development of students' reading literacy. For the rank-order component, no significant positive relations or interactions were found in the domain of reading. Altogether, these findings support the assumption that teachers' diagnostic competence in combination with instructional variables is positively related to an increase in the development of students' reading literacy. The implications of these findings for research and practice are discussed.

Theoretical Background

Teachers' diagnostic sensitivity is seen as a crucial factor for successful teaching (Helmke & Schrader, 1987; Weinert, Schrader, & Helmke, 1990). Accurate judgments of students' cognitive understanding and achievement as well as of the difficulty level of tasks and questions are required for planning and delivering instructions (Artelt & Gräsel, 2009; Helmke, Hosenfeld, & Schrader, 2004; Rogalla & Vogt, 2008; Schrader, 2011). In particular, accurate judgments are important to be able to adapt one's teaching to the students' characteristics (Helmke & Schrader, 1987; Schrader & Helmke, 2001). For example, the failure to adapt learning materials or instructions to the students' level of knowledge could lead to less learning success as well as to demotivation among students (Schrader, Helmke, Hosenfeld, Halt, & Hochweber, 2006).

Research investigating judgment accuracy usually differentiates between three different components of teacher judgment accuracy: the rank-order component, the level component, and the component of differentiation (e.g., Schrader & Helmke, 1987; Spinath, 2005; Südkamp, Möller, & Pohlmann, 2008) because measuring accuracy by only one global component results in a confusion of different judgment biases (Helmke & Schrader, 1987; Helmke et al., 2004). According to Helmke and colleagues (2004), the rank-order component is regarded as the core component of teacher judgment accuracy. This component describes the accuracy with which teachers are able to judge the rank order between their students. The rank-order component is operationalized as a correlation between teacher judgments and students' actual performance at the class level. This means that a high rank-order component is achieved if the teacher can rank his or her students in the same order as is indicated by the students' achievement on a standardized competence test. Most previous studies have shown a moderate correlation between student achievement and teacher judgments of student achievement (e.g., Hoge & Coladarci, 1989: *Mdn* r = .66; Südkamp, Kaiser, & Möller, 2012: *Mdn* r = .53). However, these studies found large variability among teachers in their judgment accuracy.

In addition to the "classic" components of teacher judgment accuracy, another component, the task-specific hit rate, can be operationalized. This component includes an exact comparison of teacher judgment and students' actual performance at the item level (see Karing, Matthäi, & Artelt, 2011). Thus, the task-specific hit rate takes into account whether teacher judgment and students' actual performance are in agreement or not. This component is based on task-specific judgments of individual students. Teachers have to compare student ability with the difficulty levels of tasks. In order to achieve a high task-specific hit rate, teachers need good knowledge about the individuals as well as good knowledge about the tasks' characteristics. This means that there is an overlap between teachers' diagnostic competence, content knowledge, and pedagogical content knowledge (Helmke, Hosenfeld, & Schrader, 2003; Karing et al., 2011). Little research has actually taken this component into account. For example, Coladarci (1986) found that elementary school teachers correctly judged 73% of their students' answers in the domain of reading. A similar result for elementary school teachers was found by Demaray and Elliott (1998), who reported that the teachers accurately gauged 79% of their students' answers in the domain of reading. Findings from the COACTIV study (Brunner, Anders, Hachfeld, & Krauss, 2011) showed a different picture of secondary school teachers. In this study, a task-specific hit rate of only 51% was obtained for mathematics teachers. However, beyond the differences between the studies in the reported mean accuracy level, all studies showed considerable differences between teachers regarding their judgment accuracy of students' performance.

Relevance of Teacher Judgment Accuracy for Students' Learning Success

According to Helmke and Schrader (1987), one prerequisite for effective teaching is the ability to adapt one's teaching to the characteristics of one's students (e.g., appropriate difficulty levels for tasks and questions, optimal instructional events). In order for teachers to adapt their teaching behavior to individual differences among students, they must have adequate diagnostic knowledge about students' abilities as well as about the difficulty levels of tasks and questions. Thus, the combination of accurate teacher judgments and adequate instructional techniques should be critical for successful teaching (Haag & Lohrmann, 2007; Helmke & Schrader, 1987; Ingenkamp, 1992; Schrader & Helmke, 2001). Despite the assumption that teacher judgments play an important role in effective teaching, it is surprising that only a few empirical studies have thus far examined the relation between teacher judgment accuracy and students' learning success. These studies have been restricted to the domain of mathematics and have shown heterogeneous results. In the study by Helmke and Schrader (1987, see also Schrader, 1989), secondary school teacher judgment accuracy was not related in general to the development of mathematical competence in lower academic track students in Grade 5. Teacher judgment accuracy was operationalized as the correlation between teachers' predicted scores for individual students and students' actual performance on a mathematics test (rank-order component). However, a significant interaction between teacher judgment accuracy and the frequency of structural cue use as well as individualized supportive contact was found: Students' learning success was highest when high judgment accuracy was combined with high instructional quality (high frequency of structural cue use or supportive individual contact). Teachers' use of structural cues included, among other things, attention-regulating comments emphasizing important information and teachers' supportive individualized contact as reflected by teachers' individual contact with their students during class work (e.g., giving students tips or answering their questions). Taken together, teacher judgment accuracy was important for students' learning success but only after taking teachers' instructional practices into account.

Lehmann and colleagues (2000) found a positive relation between teachers' accuracy in judging the difficulty levels of mathematics tasks (rank-order component) and students' mathematical competence at least for some grades and school types. A similar result was obtained by Anders and colleagues (2010; see also Brunner et al., 2011). In their study, teacher judgment accuracy was investigated using two indicators. First, the accuracy of secondary school teachers in estimating the difficulty levels of mathematics tasks in their classes (task-related bias as the mean absolute value between the actual proportion of correct answers in class and teacher judgments) and, second, their accuracy in judging the rank order of different students with regard to the students' mathematical achievement (rank-order component). The authors found a significant relation between the two indicators and the development of students' mathematical competence from Grade 9 to Grade 10. Furthermore, the relation between teachers' accuracy in judging the difficulty levels of mathematics tasks and students' learning success was mediated by teachers' cognitive activation potential in mathematics instruction. Thus, a higher accuracy in judging the difficulty levels of mathematics tasks was related to a higher cognitive activation potential, which, in turn, had a positive influence on the development of students' mathematical achievement. However, this was not found for the relation between the rank-order component and the development of students' mathematical competence.

Research Questions

As outlined above, the research that has been conducted in this area so far has focused in particular on the rank-order component, thus neglecting other measures of teacher judgment accuracy. Furthermore, studies that took the task-specific hit rate into account were restricted to elementary school teachers. Finally, little research has actually been conducted on the relation between teacher judgment accuracy and students' learning success and this research has been restricted to the domain of mathematics. Consequently, the following research questions will be addressed in this chapter:

1. How accurately can teachers judge students' reading literacy? In line with previous research (e.g., Coladarci, 1986; Demaray & Elliott, 1998; Hoge & Coladarci, 1989), we hypothesized, for both components, that teachers would exhibit a moderate to high judgment accuracy.

2. Is there a relation between teacher judgment accuracy and the development of students' reading literacy? Based on the assumption that teacher judgment accuracy is important for students' learning success (e.g., Brunner et al., 2011; Kuntze, 2006), we expected that both indicators of teacher judgment accuracy would be positively related to the development of students' reading literacy (see Figure 1, path a).

3. Is the relation between teacher judgment accuracy and the development of students' reading literacy moderated by instructional variables such as teachers' use of individualization and structural cues? In line with the findings by Helmke and Schrader (1987; see also Schrader, 1989), we predicted that the relation between teacher judgment accuracy and the development of students' reading literacy would be moderated by instructional variables (see Figure 1, path b).

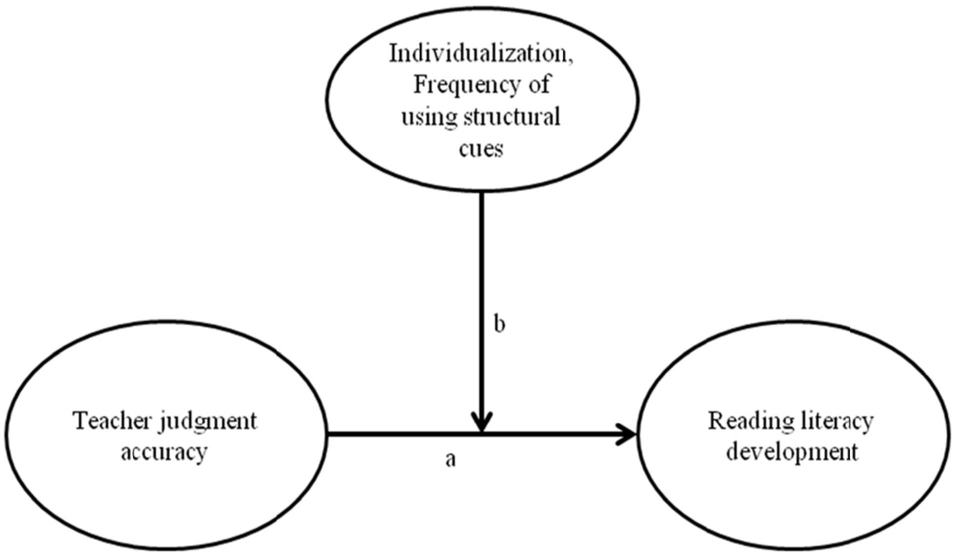

Figure 1. The expected relation between teacher judgment accuracy and the development of student reading literacy (path a), which might be moderated by several instructional variables (path b).

Method

Participants

Longitudinal data were obtained from a sample of 502 students and their 40 German language teachers at the ends of Grade 5 (T1) and Grade 6 (T2) in the context of the BiKS-8-14 study. The German language teachers were on average 46.5 years old (SD = 12.0) at T1 and had 17.8 years of teaching experience (SD = 11.1). More than half of these teachers were female (57.5 %).

The student sample consisted of 294 (58.6%) females and 208 (41.4%) males. Their mean age at T1 was 11.4 years (SD = 0.4). About 15% of the students had immigration backgrounds. The students attended 29 secondary schools (lower, middle, and higher academic tracks) across Germany (28 secondary classes in Bavaria, 12 in Hesse). Ten percent of the students were from the lower academic track ("Hauptschule"), 12.5% were from the middle academic track ("Realschule"), and 77.5% were from the higher academic track schools ("Gymnasium").

Instruments

Student variables.

Reading literacy. To assess students' reading literacy, we used sample texts with 43 multiple-choice items at the end of Grade 5 and sample texts with 31 multiple-choice items at the end of Grade 6. The reading tests were developed by the BiKS research group. For the reading literacy tests, the students had to read a given text, search the text for relevant information, and make more or less high inferences from the text to answer the given items. These tests were linked by a common item design with nonequivalent groups (anchor-item test design; see Holland, Dorans, & Petersen, 2007; Kolen & Brennan, 2004) to obtain a common metric of the individual reading literacy estimators. First, for all the reading literacy items at T1, the item difficulty parameters were estimated within an Item Response Theory framework (1-parameter Rasch model) by using the ConQuest software package (Wu, Adams, Wilson, & Haldane, 2007). Subsequently, every item difficulty parameter was fixed to guarantee a common metric of the individual reading literacy estimator. The individual student's ability was estimated by Weighted Likelihood Estimates (WLEs). In the next step, for all the reading literacy items at T2, the item difficulty parameters were estimated by using the fixed item difficulty parameters from T1 (separate estimation), and WLEs were estimated for students' ability at T2. With regard to the criterion validity, a correlation of $r = -.39$ between the reading literacy test and German grades at T1 was obtained. At T2, a correlation of $r = -.40$ was found. The internal consistency (Cronbach's alpha) of the test was satisfactory at both measurement points ($\alpha_{T1} = .79$, $\alpha_{T2} = .82$).

General cognitive abilities. At the end of Grade 5, students' general cognitive abilities were measured by a set of 15 items from the matrices subtest of the *CFT-20-R* (German version, Weiß, 2006). This test assessed the ability to recognize and solve problems of figural relations and of formal figural reasoning with different levels of complexity. The tasks contained a 2x2 or 3x3 matrix, but one cell was left blank. The student had to fill in the correct answer by choosing one out of five provided alternatives. According to the test manual, the psychometric properties of the test are acceptable (the correlation between the matrices subtest and the total test score is $r = .82$).

Table 1. Means and Standard Deviations (SD) of Students' Competence for the Total Sample and for the Different School Types in Grade 5(T1) and Grade 6 (T2)

	Total (N = 502)		Lower and middle academic tracks (N = 113)		Higher academic track (N = 389)	
	M (SD) T1	*M (SD)* T2	*M (SD)* T1	*M (SD)* T2	*M (SD)* T1	*M (SD)* T2
Reading literacy	0.91 (0.77)	1.23 (0.99)	0.31 (0.72)	0.47 (0.74)	1.08 (0.69)	1.45 (0.95)
General cognitive abilities (T1)	11.45 (2.10)	—	10.64 (2.18)	—	11.76 (1.97)	—

Note. For reading literacy, WLE scores are depicted; for general cognitive abilities, raw scores were used.

Socioeconomic status. Students' socioeconomic status was measured using the International Socio-Economic Index of Occupational Status, which is based on family members' income and educational background (ISEI; Ganzeboom, De Graaf, & Treiman, 1992). The ISEI scale ranges from 16 (low socioeconomic status) to 90 (high socioeconomic status). For the present study, we used the highest socioeconomic status in the family (HISEI). The mean HISEI at T1 of the analyzed sample was

$M = 55.8$ (SD = 16.7), whereas the national average HISEI was $M = 47.6$ in 2008 (Mikrozensus 2008; see Nold, 2010).

Teacher variables.

Teacher judgment accuracy. Teacher judgment accuracy of students' reading literacy was assessed in Grade 5 using a questionnaire that contained one reading literacy text with seven multiple-choice items and the judgment measures. This reading literacy text was chosen because of good item discrimination values and item difficulty values (with low, medium and high difficulty items). To reduce the workload for the teachers, we randomly selected seven students from each class. Teachers were asked to indicate whether each of the randomly selected students would pass (coded as 1) or fail (coded as 0) each item on the reading literacy test (see Figure 2).

Student name	Student code	Could the student answer this question correctly?						
		1	2	3	4	5	6	7
	--------------	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐

Figure 2. The judgment form on which the teacher indicated whether each of the students would pass or fail each item.

On the basis of these judgments, two indicators of teacher judgment accuracy - the rank-order component and the task-specific hit rate - were calculated. The rank-order component was computed as the correlation between teacher judgment and students' actual performance at the class level. Teacher judgment was computed by summing the number of items that the teacher had judged the student would pass (each coded as 1). A measure of each student's performance was formed by summing each student's correct answers (each correct answer was coded as 1). The task-specific hit rate was computed by summing the number of items for which a teacher's judgment and a student's actual performance were in agreement and then dividing by the number of items (see Karing et al., 2011, and the Appendix).

Teachers' use of individualization during lessons. The degree of individualization of lessons was measured in Grade 5 by a short scale consisting of four items (adapted from Clausen, 2002). The Likert-type response scale ranged from 1 (*I disagree*) to 4 (*I agree*). An example item is: "Depending on students' abilities, they are given tasks with different difficulty levels." Cronbach's alpha was $\alpha = .85$, indicating a satisfactory internal consistency.

Teachers' use of structural cues during lessons. The use of structural cues during lessons was assessed by three items in Grade 5. An example item is: "I summarize the lesson so they can remember the gist" (adapted from Rakoczy, Buff, & Lipowsky, 2005, and self-developed items). The Likert-type response scale ranged from 1 (*never*) to 4 (*very much*). Internal consistency for the scale was satisfactory, reaching $\alpha = .80$.

Table 2. Means and Standard Deviations (SD) for Instructional Variables (Individualization, Structural Cues) for the Total Sample in Grade 5 and for the Different School Types

	Total *M (SD)*	Lower and middle academic tracks *M (SD)*	Higher academic track *M (SD)*	*t*
Individualization	2.28 (0.74)	2.84 (0.82)	2.07 (0.59)	3.29*
Structural cues	3.43 (0.71)	3.79 (0.40)	3.30 (0.76)	2.01#

Note. Total: $N = 40$. Lower and middle academic tracks: $N = 11$. Higher academic track: $N = 29$.
# $p < .10$. * $p < .05$.

Statistical Analyses

To test the second and third research questions, the nested structure of the data (students are nested within classes) had to be taken into account. Multilevel analyses were applied because they integrate analyses between the student and class levels (Raudenbush & Byrk, 2002). HLM 6.08 software (Raudenbush, Bryk, Cheong, & Congdon, 2004) was used as a tool for multilevel analyses. Random intercept fixed-slope models were estimated. The method of estimation was restricted maximum likelihood. For the analyses, continuous variables were z-standardized. The variable indicating the school track was dummy-coded (0 = lower and middle academic tracks, 1 = higher academic track). The lower and middle academic tracks were combined into one category because of their small sample sizes.

The dependent variable in all models was students' reading literacy in Grade 6. First, the intercept-only model was computed to determine the intraclass correlation. Second, we specified a model that included only a set of control variables that typically affect reading literacy (Baumert & Schümer, 2001; Drechsel & Artelt, 2007; Hornberg, Valtin, Potthoff, Schwippert, & Schulz-Zander, 2007; Retelsdorf & Möller, 2008; Schaffner, Schiefele, & Schneider, 2004) and which were also used as controls on the individual level in the multilevel context in the study by Anders and colleagues (2010). Control variables at the student level (Level 1) consisted of general cognitive abilities at T1, HISEI at T1, and gender and reading literacy at T1. At the second level, the class level, school type was controlled. Third, the rank-order component and the task-specific hit rate were included separately at the second level in the model to examine the relation between teacher judgment accuracy and the development of students' reading literacy. Then, to test whether this relation was moderated by instructional variables, median splits were computed for both instructional variables (individualization: $Mdn = 2.25$; structural cues: $Mdn = 3.67$). Finally, within each subsample (low and high degree of individualization, low and high frequency of structural cue use), the influences of the rank-order component and the task-specific hit rate on reading literacy were analyzed separately while controlling for reading literacy at T1, cognitive abilities, HISEI, gender, and school type.

Missing values at the student level were imputed (m = 5) by using the multiple imputation module in the SPSS software package. All analyses were run five times, and the estimated results were automatically integrated by the HLM software.

Results

1. How Accurately do Teachers Judge Students' Reading Literacy?

German language teachers showed a mean task-specific hit rate of $M = 0.66$ ($SD = 0.11$), meaning that they correctly judged 66% of their students' answers in the domain of reading. For the rank-order component, a mean correlation of $\bar{r}$ = .19 ($SD = 0.51$) in the domain of reading was found. The results for teacher judgment accuracy are presented in Table 3. The standard deviations for the two indicators of teacher judgment accuracy of students' reading literacy indicated that there was large variability among teachers in their judgment accuracy (see Table 3 as well as Figures 3 and 4).

Table 3. Teacher Judgment Accuracy: Task-Specific Hit Rate and Rank-Order Component in Grade 5

	M	*SD*	*Min*	*Max*
Task-specific hit rate	0.66	0.11	0.40	0.86
Rank-order component	.19	0.51	-.85	.93

Note. $N = 38$ - 40 teachers. For the rank-order component, the average correlation was computed using Fisher's Z transformation.

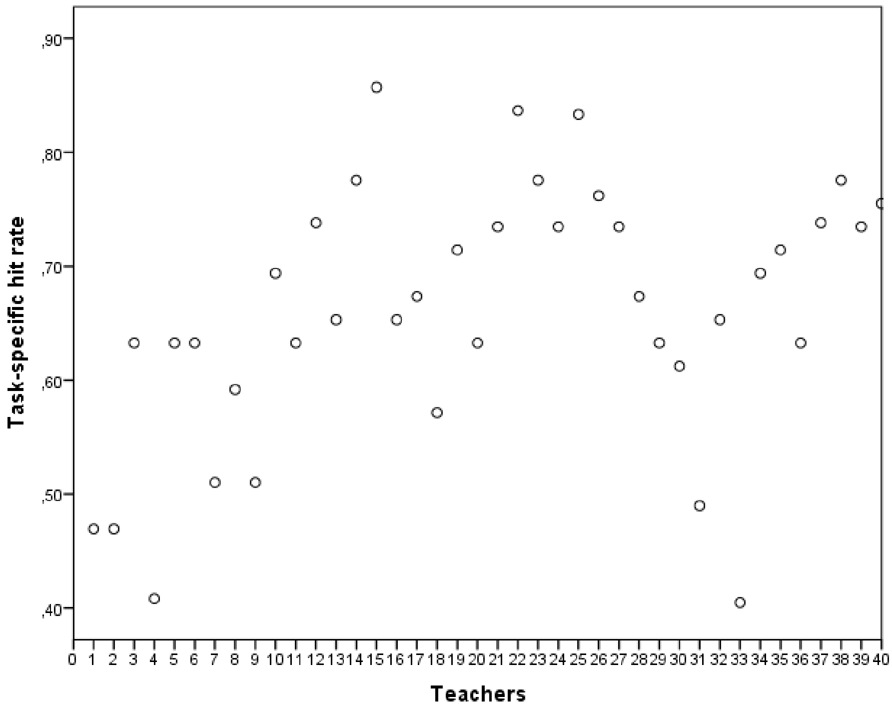

Figure 3. Teachers' judgment accuracy: task-specific hit rate.
Theoretical Range: *Min* = 0, *Max* = 1.00.

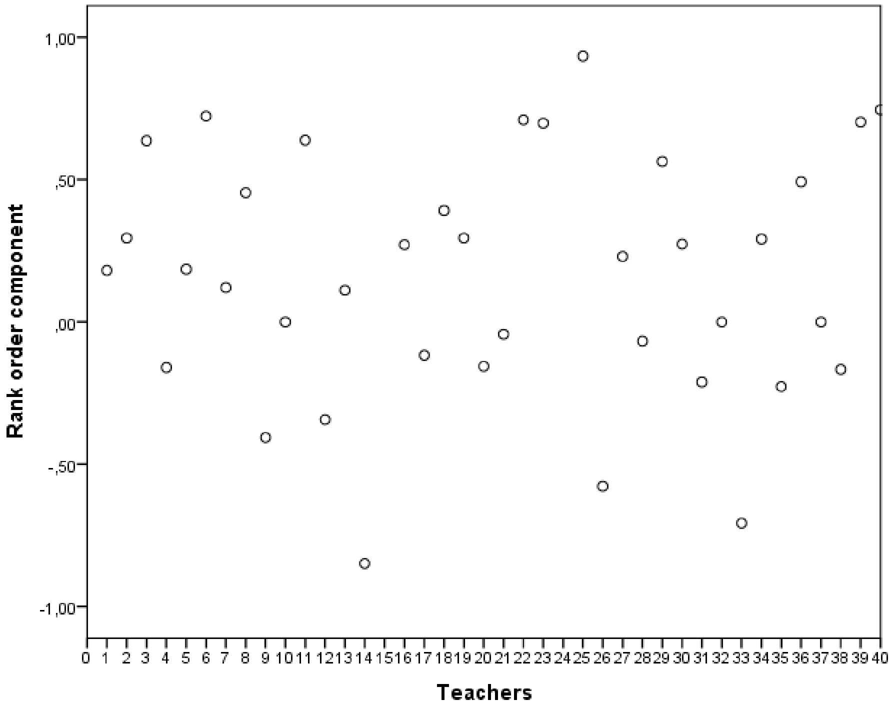

Figure 4. Teachers' judgment accuracy: rank-order component.
Theoretical Range: *Min* = -1.00, *Max* = 1.00.

2. Is There a Relation between Teacher Judgment Accuracy and the Development of Students' Reading Literacy?

Results for the second research question are presented in Table 4. First, the intercept-only model revealed an intraclass correlation coefficient of .337, meaning that 33.7% of the variance occurred between classes at T2. Consequently, 76.3% of the total variance occurred at the individual level. Second, the student- and class-level (control-) variables were included in the model (Model 1). At the student level, we found that reading literacy at T1 ($B = 0.37$, $p < .01$), gender ($B = 0.29$, $p < .01$), general cognitive abilities ($B = 0.13$, $p < .01$), and HISEI ($B = 0.11$, $p < .05$) were significantly related to students' reading literacy at T2. Moreover, a significant effect of school type at the class level ($B = 0.37$, $p < .01$) was found. Furthermore, results indicated that 40.2% of the total variance was explained by the student- and class-level variables. Third, the task-specific hit rate (Model 2) and the rank-order component (Model 3) were introduced separately into the model to examine the relation between teacher judgment accuracy and the development of students' reading literacy. Analyses revealed that the task-specific hit rate had a significant positive relation with the development of students' reading literacy ($B = 0.15$, $p < .05$, see Model 2),[2] whereas the rank-order component was not significantly positively related to the development of students' reading literacy ($B = -0.02$, $p > .05$, see Model 3). The percentage of total variance that was explained by Model 2 was 41.9%; by Model 3, it was 40.7%.

Table 4. Results from the Multilevel Analyses Predicting Reading Literacy in Grade 6 (T2)

	Intercept-only model	Model 1		Model 2		Model 3[b]	
		B	*SE (B)*	*B*	*SE (B)*	*B*	*SE (B)*
Class level							
School type: higher track[a]		0.37**	0.11	0.19	0.18	0.37**	0.11
Task-specific hit rate (T1)		—	—	0.15*	0.09	—	—
Rank-order component (T1)				—	—	-0.02	0.05
Student level							
Reading literacy (T1)		0.37**	0.04	0.36**	0.04	0.36**	0.04
General cognitive abilities (T1)		0.13**	0.03	0.13**	0.03	0.13**	0.03
HISEI (T1)		0.11*	0.05	0.10*	0.05	0.10*	0.05
Gender		0.29**	0.07	0.29**	0.06	0.27**	0.06
Intercept (γ_{00})		0.77**	0.08	0.90**	0.11	0.77**	0.09
Residual variance							
Class level (u_0)	0.337	0.081		0.065		0.090	
Student level (r)	0.663	0.517		0.516		0.467	
ICC	0.337	0.135		0.112		0.162	
R^2		0.402		0.419		0.407	

Note. [a] reference: lower and middle academic tracks; gender: 0 = male, 1 = female; five classes in the lower academic track, six classes in the middle academic track, and 29 classes in the higher academic track; 502 students; [b] Model 3: five classes in the lower academic track, six classes in the middle academic track, and 27 classes in the higher academic track; 476 students; ICC: intraclass correlations (variance between classes [u_0]/ total variance [$r + u_0$]).
**p* < .05. ***p* < .01.

[2] Higher academic track teachers had a significantly higher task-specific hit rate than teachers in the middle and lower academic tracks ($t = 3.92$, $p < .01$; higher academic track: $M = 0.70$, $SD = 0.10$; middle and lower academic tracks: $M = 0.56$, $SD = 0.09$). Thus, the analyses were computed again using only the higher academic track teachers. The results show a positive relation between the task-specific hit rate and the development of reading literacy but reached statistical significance only at the 10% level ($B = 0.17$, $p < .10$).

3. Is the Relation between Teacher Judgment Accuracy and the Development of Students' Reading Literacy Moderated by Instructional Variables?

To test whether this relation was moderated by instructional variables, multilevel analyses were computed separately for each subsample (low and high degree of individualization, low and high frequency of structural cue use during lessons). First, the results for teachers' use of individualization during lessons are presented (Table 5). For teachers who used a high degree of individualization during lessons, we found a significant positive relation between the task-specific hit rate and the development of students' reading literacy (B = 0.23, $p < .05$), whereas for teachers who applied a low degree of individualization during lessons, the task-specific hit rate was not significantly related to the development of students' reading literacy (B = 0.10, $p > .05$). For the rank-order component, again, no relation between this indicator and the development of students' reading literacy was found: A significant positive relation between the rank-order component and students' reading literacy development was not demonstrated in the group with a low degree of individualization (B = -0.04, p > .05) or in the group with a high degree of individualization (B = 0.03, p > .05).

Table 5. Results from the Multilevel Analyses Predicting Reading Literacy in Grade 6 (T2) Separately for Low and High Degrees of Individualization

	Low degree of individualization				High degree of individualization			
	Model 1		Model 2[b]		Model 1		Model 2	
	B	*SE (B)*	*B*	*SE (B)*	*B*	*SE (B)*	*B*	*SE (B)*
Class level								
School type: higher track[a]	0.39	0.19	0.50**	0.16	-0.02	0.21	0.31*	0.13
Task-specific hit rate (T1)	0.10	0.09	—	—	0.23*	0.09	—	—
Rank-order component (T1)	—	—	-0.04	0.04	—	—	0.03	0.11
Student level								
Reading literacy (T1)	0.42**	0.08	0.41**	0.08	0.30**	0.04	0.32**	0.04
General cognitive abilities (T1)	0.16**	0.05	0.19**	0.05	0.11**	0.04	0.10**	0.04
HISEI (T1)	0.00	0.09	-0.03	0.08	0.17**	0.04	0.18**	0.05
Gender	0.28*	0.11	0.23	0.12	0.28**	0.07	0.30**	0.07
Intercept (γ_{00})	0.87**	0.10	0.81**	0.13	0.96**	0.16	0.72**	0.11
Residual variance								
Class level (u_0)	0.054		0.083		0.047		0.081	
Student level (r)	0.612		0.503		0.426		0.427	
ICC	0.081		0.142		0.099		0.159	
R^2	0.361		0.379		0.467		0.427	

Note. [a] reference: lower and middle academic tracks; gender: 0 = male, 1 = female; low degree of individualization: four classes in the middle academic track and 14 classes in the higher academic track; 235 students; [b] Model 2: four classes in the middle academic track and 12 classes in the higher academic track; 209 students; high degree of individualization: five classes in the lower academic track, two classes in the middle academic track, and 15 classes in the higher academic track; 267 students; ICC: intraclass correlations (variance between classes [u_0]/ total variance [$r + u_0$]).
$^*p < .05$. $^{**}p < .01$.

With regard to the teachers' use of structural cues during lessons (Table 6), an unexpected result was found. The results revealed that for teachers with a low frequency of structural cue use during lessons, there was a significant positive relation between the task-specific hit rate and the development of students' reading literacy ($B = 0.36$, $p < .01$), whereas no significant relation was found for teachers who frequently used structural cues during lessons ($B = 0.01$, $p > .05$).[3] Again, no significant relation between the rank-order component and the development of students' reading literacy was found, either in the subsample with a low frequency of

[3] Analyses were computed again using only the higher academic track teachers because of their significantly higher task-specific hit rate compared to teachers in the middle and lower academic tracks. The findings again showed a positive relation between the task-specific hit rate and the development of reading literacy for teachers with a high degree of individualization ($B = 0.35$, $p < .01$) and for teachers who used few structural cues during lessons ($B = 0.36$, $p < .01$). For the other groups, no significant relations were found.

structural cues ($B = 0.05$, $p > .05$) or in the subsample with frequently used structural cues ($B = -0.07$, $p > .05$).

Table 6. Results from the Multilevel Analyses Predicting Reading Literacy in Grade 6 (T2) Separately for Low and High Frequencies of Structural Cue Use During Lessons

	Low frequency of structural cue use				High frequency of structural cue use			
	Model 1		Model 2[b]		Model 1		Model 2[c]	
	B	*SE (B)*	*B*	*SE (B)*	*B*	*SE (B)*	*B*	*SE (B)*
Class level								
School type: higher track[a]	-0.14	0.22	0.33	0.19	0.36*	0.17	0.38*	0.15
Task-specific hit rate (T1)	0.36**	0.09	—	—	0.01	0.07	—	—
Rank-order component (T1)	—	—	0.05	0.08	—	—	-0.07	0.05
Student level								
Reading literacy (T1)	0.36**	0.08	0.34**	0.08	0.36**	0.05	0.36**	0.05
General cognitive abilities (T1)	0.11*	0.05	0.12*	0.05	0.15**	0.03	0.14**	0.04
HISEI (T1)	0.06	0.06	0.07	0.06	0.15*	0.06	0.12	0.07
Gender	0.20*	0.08	0.21	0.08	0.35**	0.09	0.31**	0.09
Intercept (γ_{00})	1.14**	0.19	0.80**	0.17	0.78**	0.09	0.77**	0.10
Residual variance								
Class level (u_0)	0.046		0.165		0.041		0.047	
Student level (r)	0.559		0.502		0.489		0.446	
ICC	0.076		0.248		0.077		0.095	
R^2	0.381		0.261		0.476		0.491	

Note. [a] reference: lower and middle academic tracks; gender: 0 = male, 1 = female; few structural cues: two classes in the middle academic track and 15 classes in the higher academic track; 213 students; [b] Model 2: two classes in the middle academic track and 14 classes in the higher academic track; 204 students; a lot of structural cues: five classes in the lower academic track, four classes in the middle academic track, and 14 classes in the higher academic track; 289 students; [c] Model 2: five classes in the lower academic track, four classes in the middle academic track, and 13 classes in the higher academic track; 272 students; ICC: intraclass correlations (variance between classes [u_0]/ total variance [$r + u_0$]).
$^*p < .05$. $^{**}p < .01$.

Discussion

The aim of the study was, on the one hand, to examine the accuracy of secondary school teacher judgments in the domain of reading and, on the other hand, to investigate the relation between teacher judgment accuracy and the development of students' reading literacy and whether this relation was moderated by instructional variables.

As predicted, German language teachers showed a moderate hit rate. They accurately judged 66% of their students' answers. Compared to findings from studies with elementary school teachers, the task-specific hit rate in our study was smaller than the task-specific hit rate reported in studies with elementary school teachers (see Coladarci, 1986; Demaray & Elliott, 1998). For the rank-order component, we found quite a low correlation between teacher judgments and students' actual reading performance ($\bar{r}$ = .19). A possible explanation for the low correlation could be that the seven items of the reading literacy test did not differentiate sufficiently well between the students. Such a restriction of variance can substantially depress correlations and thus lead to an understatement of the relation between teacher judgments and students' actual performance. In addition, the rank-order component in our study was also smaller than the rank-order component reported in studies with elementary school teachers (Demaray & Elliott, 1998; Feinberg & Shapiro, 1998). For example, Demaray and Elliott (1998) obtained a correlation of r = .82 in the domain of reading. However, the elementary school teachers in their study had to rate each item (52 items) on the reading literacy test, whereas the secondary school teachers in our study had to judge only seven items on the reading literacy test, thus resulting in a restriction of variance. Furthermore, there were differences in the manner in which the correlations were operationalized. In the present study, mean within-class correlations were computed, whereas in the study by Demaray and Elliott (1998), class membership was not considered when computing correlations. Looking only at the overall correlation and ignoring the class level, as done by Demaray and Elliot (1998), leads to a confounding of differences between classes and differences between students within classes and may therefore be affected by substantial bias (Schrader & Helmke, 1990).

Another possible explanation could be that actual differences between elementary and secondary school teachers exist. The first evidence for this comes from a study by Karing (2009), who found that elementary school teachers more accurately judged students' reading literacy as well as students' mathematical competence than secondary school teachers. This finding is consistent with the assumption that differences between elementary and secondary school teacher judgment accuracy are related to structural prerequisites such as class composition (e.g., heterogeneity of students' achievement) and teachers' education. For example, elementary school classes are more heterogeneous with regard to the academic performance of students than secondary school classes (Tillman & Wischer, 2006). Some studies have found that the heterogeneity of student achievement was positive related to measures of teachers' diagnostic competence (Schrader, 1989; Weinert & Lingelbach, 1995; Karing, 2009). Furthermore, the elementary school teachers had more pedagogical and psychological practice in their teacher training than secondary school teachers, especially teachers in the higher academic track (Einsiedler, 2004; Hermann, 2004).

As in previous studies (e.g., Hoge & Coladarci, 1989; Südkamp et al., 2012), we found large variability among German language teachers in their judgment accuracy of students' reading literacy. The variability in correlations ranged from -.85 to .92, and the range for the task-specific hit rate ranged from 40% to 86%, meaning that there may be a substantial number of different variables (e.g., student and teacher characteristics) that influence teacher judgment accuracy of students' academic performance.

With regard to our analyses of the relation between teacher judgment accuracy and the development of students' reading literacy, the following results were found: First, as expected, a significant positive relation between teachers' task-specific hit rate and the development of students' reading literacy was demonstrated. However, the significant relation between teachers' task-specific hit rate and the development of students' reading literacy was moderated by instructional variables: A high task-specific hit rate in combination with a high degree of individualization of lessons was significantly associated with an increased development of students' reading literacy, whereas a high task-specific hit rate in combination with a low degree of individualization of lessons had no effect on students' reading literacy development. Furthermore, a high task-specific hit rate in combination with a low frequency of structural cue use during lessons was also significantly related to an increase in the development of students' reading literacy. However, no relation was demonstrated when structural cues were frequently used. A possible explanation for this unexpected finding may be that high-ability students do not depend on teachers' use of structural cues during lessons, but rather rely on self-directed learning and individualized instructions because of their favorable learning prerequisites. On the other hand, for low-ability students, a highly structured learning environment makes it easier for them to focus their attention on relevant aspects of the lessons and to more easily combine prior knowledge with new knowledge (Blumberg, Möller, & Hardy, 2004; Lipowsky, 2009). For example, Möller, Jonen, Hardy, and Stern (2002) found that high-ability elementary school students did not require highly structured lessons in social studies and science for their learning success, whereas low-ability students profited more from a highly structured learning environment. This could explain the difference between the findings of Helmke and Schrader (1987; see also Schrader 1989) and our study. Contrary to our study, Helmke and Schrader (1987; see also Schrader 1989) found that mathematics achievement gains were highest when high diagnostic competence was combined with the use of a lot of structural cues during lessons. In our study however, nearly 80% of the students attended higher academic track schools, whereas in the study by Helmke and Schrader (1987; see also Schrader 1989), the sample consisted exclusively of lower academic track students.

With regard to the rank-order component, no significant positive relation with the development of students' reading literacy was found. Furthermore, we found no significant interaction between this indicator and either instructional variable for the development of reading literacy. One reason for the different findings regarding the association between the two indicators of teacher judgment accuracy and the development of reading literacy might be the low correspondence between teacher judgment accuracy and students' actual performance. Perhaps as a result of the low value of the rank-order component, no significant relation with students' development of reading literacy could be identified. According to Schrader (1989), a minimal degree of diagnostic competence as well as instructional quality is necessary to achieve significant relations or interactions. Maybe there were not enough teachers in our study who showed the necessary minimal degree of this indicator (rank-order component) to achieve significant relations and interactions.

Limitations and Suggestions for Future Research

Our study has some limitations that need to be taken into account. A perfect correspondence between teacher judgment of students' performance and students' actual performance cannot be expected because the competence tests reflect only a single performance of the students, whereas teacher judgments within the school context are based on different oral and written performances of the students. As in previous studies, the reliabilities of the two indicators of teacher judgment accuracy could not be computed (McElvany et al., 2009; Schrader, 1989). According to Schrader (2009), the reliabilities of these indicators depend on the reliability of the judgment as well as on the reliability of the criterion, but are not definitively determined by these two variables. A further limitation is that both instructional variables were based on self-reports from teachers. Thus, they are limited to the views of the teachers and may be affected by judgment biases. Furthermore, median splits were computed for both instructional variables to answer the third question. However, a consequence of dichotomization is the loss of information about individual differences as well as the loss of statistical power (MacCallum, Zhang, Preacher, & Rucker, 2002; Richter, 2007; Schrader, 1989). Finally, due to the small sample sizes of students and teachers from lower and middle academic track schools, a generalization of the present findings to these school types should be made with caution. According to Mass and Hox (2005), a sample size of at least $N = 50$ classes (Level 2) is needed for multilevel analyses. In our study, we had only $N = 40$ classes, leading to low test power and high insecurity in the estimation of the model parameters.

Despite these limitations, this study makes an important contribution to our understanding of the relation between teacher judgment accuracy and the development of students' reading literacy. It shows that a combination of both high diagnostic sensitivity and appropriate instructional practices by teachers is necessary for effective teaching. However, our study, like previous studies (e.g., Anders et al., 2010; Helmke & Schrader, 1987; Lehmann et al., 2000), investigated the relation between teacher judgment accuracy and students' learning success only in secondary school and only in the domains of reading and mathematics. Thus, further research is needed to take into account elementary school teachers as well as other domains and school subjects.

Results from the present study, like previous work (e.g., Anders et al., 2010; McElvany et al., 2009; Schrader, 1989), showed general deficits in teacher judgment accuracy. Along with the assumption that teacher judgment accuracy is important for successful teaching, there is a considerable need for special teacher training. A first approach is offered by VERA ("Vergleichsarbeiten"; Helmke et al., 2004). Here, elementary school teachers get feedback about their judgment accuracy in the domains of reading and mathematics (task-related rank-order component and level component). However, merely informing teachers about the accuracy of their judgments does not seem to be sufficient for substantially improving their judgment accuracy. Rather, improving their judgment accuracy depends on how teachers apply this information about their judgment accuracy. Unfortunately, not much is known about this important topic, which urgently needs further exploration.

References

Anders, Y., Kunter, M., Brunner, M., Krauss, S., & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht, 57*, 175-193. doi 10.237/peu2010.arT13d.

Artelt, C., & Gräsel, C. (2009). Gasteditorial. Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie, 23*, 157-160. doi: 10.1024/1010-0652.23.34.157.

Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, J. M. Schneider et al. (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323-407). Opladen: Leske & Budrich.

Blumberg, E., Möller, K., & Hardy, I. (2004). Erreichen motivationaler und selbstbezogener Zielsetzungen in einem schülerorientierten naturwissenschaftsbezogenen Sachunterricht - Bestehen Unterschiede in Abhängigkeit von der Leistungsstärke?. In W. Bos, E. Lankes, N. Plaßmeier, & K. Schwippert (Hrsg.), *Heterogenität. Eine Herausforderung an die empirische Bildungsforschung*. Münster: Waxmann.

Brunner, M., Anders, Y., Hachfeld, A., & Krauss, S. (2011). Diagnostische Fähigkeiten von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 215-234). Münster: Waxmann.

Clausen, M. (2002). *Qualität von Unterricht – Eine Frage der Perspektive?* Münster: Waxmann.

Coladarci, T. (1986). The accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology, 78,* 141-146.

Demaray, M. K., & Elliott, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly, 13*, 8-24.

Drechsel, B., & Artelt, C. (2007). Lesekompetenz. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 225-248). Münster: Waxmann.

Einsiedler,W. (2004). Lehrerausbildung für Grundschule. In Blömeke, S., Reinhold, P., Tulodziecki, G. & Wildt, J. (Hrgs.), *Handbuch Lehrerausbildung* (S. 315-324). Bad Heilbrunn: Klinkhardt.

Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-basedjudgments of students' reading with differing achievement levels. *The Journal of Educational Research, 102,* 453-462. doi: 10.1521/scpq.18.1.52.20876.

Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research 21*, 1-56.

Haag, L., & Lohrmann, K. (2007). Diagnostische (In-)Kompetenz von Lehrern. In B. Fuchs & C. Schönherr (Hrsg.), *Urteilskraft und Pädagogik. Beiträge zu einer pädagogischen Handlungstheorie*. Würzburg: Königshausen & Neumann.

Helmke, A., Hosenfeld, I., & Schrader, F. W. (2003). Diagnosekompetenz in Ausbildung und Beruf entwickeln. *Karlsruher pädagogische Beiträge, 55*, 15-34.

Helmke, A., Hosenfeld, I., & Schrader, F. W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119-144). Hohengehren: Schneider Verlag.

Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education, 3*, 91-98.

Herrmann, U. (2004). Lehrerausbildung für das Gymnasium und die Gesamtschule. In Blömeke, S., Reinhold, P., Tulodziecki, G. & Wildt, J. (Hrsg.), *Handbuch Lehrerausbildung* (S. 335-350). Bad Heilbrunn: Klinkhardt.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Psychology, 59*, 297-313.

Holland, P. W., Dorans, N. J., & Peterson, N. S. (2007). Equating test scores. In C. R. Rao & S. Sinharay (Hrsg.), *Handbook of Statistics 26. Psychometrics* (S. 169-203). Amsterdam: Elsevier.

Hornberg, S., Valtin, R., Potthoff, B., Schwippert, K., & Schulz-Zander, R. (2007). Lesekompetenzen von Mädchen und Jungen im internationalen Vergleich. In W. Bos, S. Hornberg, R. Valtin, K. Schwippert, E.-M. Lankes, L. Fried, K.-H. et al. (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 195-223). Münster: Waxmann.

Ingenkamp, K. (1992). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.

Institut für Qualitätsentwicklung (2007). *Leseverständnistest 7 Hessen*. Wiesbaden: Institut für Qualitätsentwicklung.

Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie, 23*, 197-209. doi: 10.1024/1010-0652.23.34.197.

Karing, C., Matthäi, J., & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I - Eine Frage der Spezifität?. *Zeitschrift für Pädagogische Psychologie, 25*, 159-172. doi: 10.1024/1010-0652/a000041.

Karing, C., Pfost, M., & Artelt, C. (2011). Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen?. *Journal for Educational Research Online, 3*, 119-147.

Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking. Methods and Practices.* New York, NY: Springer.

Kuntze, S. (2006). Konzeption einer Lernumgebung für den gymnasialen Mathematikunterricht und Evaluation einer Themenstudienarbeit zum mathematischen Beweisen und Argumentieren. Zugriff am 30.04.2013. Verfügbar unter: http://edoc.ub.uni-muenchen.de/6278.

Lehmann, R. H., Peek, R., Gänsfuß, R., Lutkat, S., Mücke, S., & Barth, I. (2000). *Qualitätsuntersuchungen an Schulen zum Unterricht in Mathematik (QuaSUM).* Potsdam: Ministerium für Bildung, Jugend und Sport des Landes Brandenburg (MBJS).

Lipowsky, F. (2009). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie [Educational Psychology].* (S. 73-102). Berlin: Springer.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19-40.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1,* 85-91.

McElvany, N., Schroeder, S., Richter, T., Hachfeld, A., Baumert, J., Schnotz, W. et al. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie, 23,* 223-235. doi: 10.1024/1010-0652.23.34.223.

Möller, K., Jonen, A., Hardy, I., & Stern, E. (2002). Die Förderung von naturwissenschaftlichem Verständnis bei Grundschulkindern durch Strukturierung der Lernumgebung. *Zeitschrift für Pädagogik. 45.Beiheft. Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen,* 176-191.

Nold, D. (2010). *Sozioökonomischer Status von Schülerinnen und Schülern 2008. Ergebnisse des Mikrozensus. Statistisches Bundesamt*. Zugriff am 30.04.2013. Verfügbar unter: https://www.destatis.de/DE/Publikationen/Wirtschaft/Statistik/BildungForschungKultur/ StatusSchueler_22010.pdf?__blob=publicationFile.

Rakoczy, K., Buff, A., & Lipowsky, F. (2005). Teil 1: Befragungsinstrumente. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung / Deutsches Institut für Internationale Pädagogische Forschung.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks: Sage.

Raudenbush, S. W., Bryk, A., Cheong, Y. F., & Congdon, R. (2004). HLM 6: Hierarchical linear and nonlinear modeling. Chicago: Scientific Software International.

Retelsdorf, J., & Möller, J. (2008). Entwicklungen von Lesekompetenz und Lesemotivation. Schereneffekte in der Sekundarstufe?. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 40,* 179-188. doi: 10.1026/0049-8637.40.4.179.

Richter, T. (2007). Wie analysiert man Interaktionen von metrischen und kategorialen Prädiktoren? Nicht mit Median-Splits! *Zeitschrift für Medienpsychologie, 19,* 116-125. doi: 10.1026/1617-6383.19.3.116.

Rogalla, M., & Vogt, F. (2008). Förderung adaptiver Lehrkompetenz: eine Interventionsstudie. *Unterrichtswissenschaft, 36,* 17-36.

Schaffner, E., Schiefele, U., & Schneider, W. (2004). Ein erweitertes Verständnis der Lesekompetenz: Die Ergebnisse des nationalen Ergänzungstests. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 197-242). Wiesbaden: VS Verlag für Sozialwissenschaften.

Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt am Main: Peter Lang.

Schrader, F.-W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie, 23*, 237-245. doi: 10.1024/1010-0652.23.34.237.

Schrader, F.-W. (2011). Lehrer als Diagnostiker. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch zur Forschung zum Lehrerberuf* [(S. 683-698). Münster: Waxmann.

Schrader, F.-W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik, 1*, 27-52.

Schrader, F.-W., & Helmke, A. (1990). Lassen sich Lehrer bei der Leistungsbeurteilung von sachfremden Gesichtspunkten leiten? Eine Untersuchung zu Determinanten diagnostischer Lehrerurteile. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 22*, 312-324.

Schrader, F.-W., & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen*. (S. 45-58). Weinheim: Beltz.

Schrader, F.-W., Helmke, A., Hosenfeld, I., Halt, A.C., & Hochweber, J. (2006). Komponenten der Diagnosegenauigkeit von Lehrkräften: Ergebnisse aus Vergleichsarbeiten. In F. Eder, A. Gastager & F. Hofmann (Hrsg.), *Qualität durch Standards? Beiträge zur 67. AEPF-Tagung in Salzburg* (S. 265-278). Münster: Waxmann.

Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie, 19*, 85-95. doi: 10.1024/1010-0652.19.1.85.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A Meta-Analysis. *Journal of Educational Psychology, 104*, 743-762. doi: 10.1037/a0027627

Südkamp, A., Möller, J., & Pohlmann, B. (2008). Der Simulierte Klassenraum. Eine experimentelle Untersuchung zur diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie, 22,* 261-276. doi: 10.1024/1010-0652.22.34.261.

Tillmann, K.-J., & Wischer, B. (2006). Heterogenität in der Schule. Forschungsstand und Konsequenzen. *Pädagogik, 3,* 44-48.

Weinert, F. E., & Lingelbach, H. (1995). Teaching Expertise: Theoretical conceptualizations, empirical findings, and some consequences for teacher training. In Hoz, R. & Silberstein, M. (Hrsg.), *Partnerships of schools and institutions of higher education in teacher development* (S. 293-302). Beer-Shera, Israel: Ben-Gurion University of the Negev Press.

Weinert, F. E., Schrader, F.-W., & Helmke, A. (1990). Educational expertise: Closing the gap between educational research and classroom practice. *School Psychology International, 11*, 163-180.

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 - Revision (CFT 20-R)*. Göttingen: Hogrefe.

Wu, M. L., Adams, R. J., Wilson, M., & Haldane, S. A. (2007). *ACER ConQuest version 2.0: generalised item response modelling software*. Camberwell: ACER Press.

Appendix

An example of the computation of the two indicators of teacher judgment accuracy. In the following table, a fictitious class is illustrated. The first part of the table contains students' achievement (passed = coded as 1; failed = coded as 0). The second part of the table includes teacher judgments. The teacher judged whether each of the seven students would pass (coded as 1) or fail (coded as 0) each of the seven items.

	Students' achievement								Teacher judgment (one teacher)								Hit rate (t_j)
	1	2	3	4	5	6	7	$\sum_{k=1}^{m} S_{ijk}$	1	2	3	4	5	6	7	$\sum_{k=1}^{m} L_{ijk}$	
S_1	1	1	1	1	1	1	1	7	0	0	1	0	1	1	1	4	4
S_2	1	1	0	0	0	0	0	2	1	1	1	1	1	1	1	7	2
S_3	0	0	1	0	1	0	0	2	0	0	0	0	0	0	0	0	5
S_4	1	0	1	0	1	1	1	5	1	1	1	1	0	1	1	6	4
S_5	0	1	0	1	0	1	1	4	1	0	1	0	1	1	1	5	2
S_6	0	0	1	0	0	0	0	1	1	0	1	1	0	0	0	3	5
S_7	1	1	1	1	1	1	0	6	1	1	1	1	1	1	1	7	6
								$SD_x = 2.27$								$SD_y = 2.51$	$\sum_{i=1}^{l} t_{ij} = 28$

Note. S_1 to S_7 = students; 1 .. 7 = items; i =1 .. l = number of students; j = 1 .. n number of teachers; k = 1 .. m number of tasks; SD_x = standard deviation of students' achievement, SD_y = standard deviation of teacher judgments, Covariance is $COV_{xy} = 2.22$,

$\sum_{j=1}^{m} \mathrm{S}_{ijk}$ = the sum of an individual student's correct answers (= individual student's performance),

$\sum_{j=1}^{m} \mathrm{L}_{ijk}$ = the sum of teacher judgements of individual students' correct answers.

Computation of two indicators of teacher judgment accuracy:

1.) Rank-order component (r_{xy}):

$$r_{xy} = \frac{\mathrm{COV_{xy}}}{\mathrm{SD_x} \times \mathrm{SD_y}} = \frac{2.22}{2.27 \times 2.51} = 0.39$$

A moderate correlation between teacher judgment and students' performance.

2.) Task-specific hit rate (aT_j):

$$aT_j = \frac{1}{m} \sum_{i=1}^{l} \mathrm{t}_{ij} = \frac{28}{49} = 0.57$$

The teacher correctly judged 57% of their students' answers.

10 Authors

Cordula Artelt, Dr. phil., is full professor (Chair of Educational Research) at the University of Bamberg, Germany. Her research interests are related to reading and text comprehension, metacognition and self-regulated learning (development and training), teacher competencies as well methods of large scale assessments and the assessment of student competencies.

Tobias Dörfler, Dr. phil., is a full professor for psychological assessment and head of the Department of Psychology at the University of Education Heidelberg, Germany. His research interests focus on individual differences in reading comprehension and the assessment of reading abilities as well as methodological foundations in longitudinal measurement.

Susanne Ebert, Dr. phil., is an assistant professor at the Department of Developmental Psychology at the University of Bamberg, Germany. Her research interest is focused on individual differences in (meta)cognitive and language development. A specific interest of her research is on children's understanding of the mental world (mental state language, metamemory, theory of mind).

Constance Karing, Dr. phil., is research scientist at the Department of Research Synthesis, Intervention and Evaluation at the University of Jena, Germany. Her research interests focus on teacher judgement accuracy and stressprevention.

Susanne Kuger, Dr. phil., is a research scientist at the German Institute for International Educational Research in Frankfurt, Germany. Her research focuses on the quality and quantity of learning opportunities as well as differential educational effectiveness.

Simone Lehrl, Dipl.-Paed., is a research scientist at the Department of Early Childhood Education at the University of Bamberg, Germany. Her research interests focus on early childhood education, especially the dynamics of the early home learning environment and its effects on children's development.

Christian Lorenz, Dr. phil., is a research scientist at the University of Bamberg, Germany. He was coordinator of the BiKS longitudinal study and is now working at the National Educational Panel Study (NEPS). His research focus is on teachers' judgment accuracy, in the field of competence development in preschool and school age and on methodological issues concerning empirical field research.

Michael Mudiappa, Dipl. Soz.wiss., is a research scientist at the University of Bamberg, Germany. He was coordinator of the BiKS longitudinal study and is now working at the National Educational Panel Study (NEPS). His research interests are longitudinal analyses in the field of educational inequality. Especially, he focuses on the influence of cultural capital on school success in primary and secondary schools.

Maximilian Pfost, Dr. phil., is an assistant professor at the Department of Educational Research at the University of Bamberg, Germany. His research interests focus on individual differences in reading comprehension and the role of schools and families for literacy development.

Hans-Guenther Rossbach, Dr. phil., is full professor (Chair of Early Childhood Education) at the University of Bamberg and primary investigator of the National Educational Panel Study (NEPS), Germany. His main research interests are the effects of early nonmaternal care on social and cognitive development and longitudinal educational research.

Monja Schmitt, Dr. rer. pol., is a research scientist at the University of Bamberg, Germany. She is coordinator of the BiKS longitudinal study and has also been working in the National Educational Panel Study (NEPS). Her research focus is on the relation between social capital and educational outcomes across the life course. Furthermore, she is working on methodological issues concerning empirical field research.

Thorsten Schneider, Dr. phil., is professor of sociology with a special focus on comparative analysis of contemporary societies at the Universität Leipzig, Germany. His research interests are in the fields of social and ethnic inequalities in educational opportunities, intergenerational transfers, and longitudinal research methods.

Irene M. Schurtz, Dipl.-Soz., is a research scientist at the Department of Educational Research at the University of Bamberg, Germany. Her research interests focus on the development of students' interests as well as students' academic self-concepts.

Sabine Weinert, Dr. phil., is full professor (Chair of Developmental Psychology) at the University of Bamberg, Germany. Her research is focused on developmental and educational psychology and specifically on the development of language, cognition, and learning including issues of developmental disorders, developmental diagnosis, promotion and intervention.

University
of Bamberg
Press

The BiKS research group ("Educational processes, competence development, and selection decisions in preschool- and school age") founded in 2005 and financed by the German Research Foundation (DFG), was established by a consortium of researchers combining expertise from the disciplines of psychology, education, and sociology. Two longitudinal studies were being conducted by the BiKS research group and followed until 2012. In the first study, a cohort of preschool children was traced until grade 4 in primary school. The second study comprises a cohort of primary school children who were followed until their 9th grade in secondary school. Besides the multidisciplinary perspective, the studies can be well characterized by their broad use of different methods, such as test data, interviews, questionnaires, and live observations of behaviour as well as a consideration of different agents, i.e. students, parents, and teachers. The book focuses on empirical research findings concerning the development of reading literacy from a longitudinal perspective and the chapters cover findings from both longitudinal studies of the BiKS research group. As authors from different academic disciplines have contributed, this volume covers a range of psychological, educational as well as sociological perspectives on causes and effects of stability and interindividual differences in the development of reading literacy.

eISBN 978-3-86309-165-1

9 783863 091651

www.uni-bamberg.de/ubp/