

RESEARCH

Open Access



# Participation rates, characteristics, and differential effects on reading literacy of extracurricular tutoring in a German large-scale assessment

Nora Heyne<sup>1\*</sup> , Timo Gnamb<sup>2</sup>  and Marie-Ann Sengewald<sup>2</sup> 

\*Correspondence:  
[nora.heyne@uni-bamberg.de](mailto:nora.heyne@uni-bamberg.de)

<sup>1</sup> Department of Educational Research, University of Bamberg, Markusstraße 8a, 96047 Bamberg, Germany

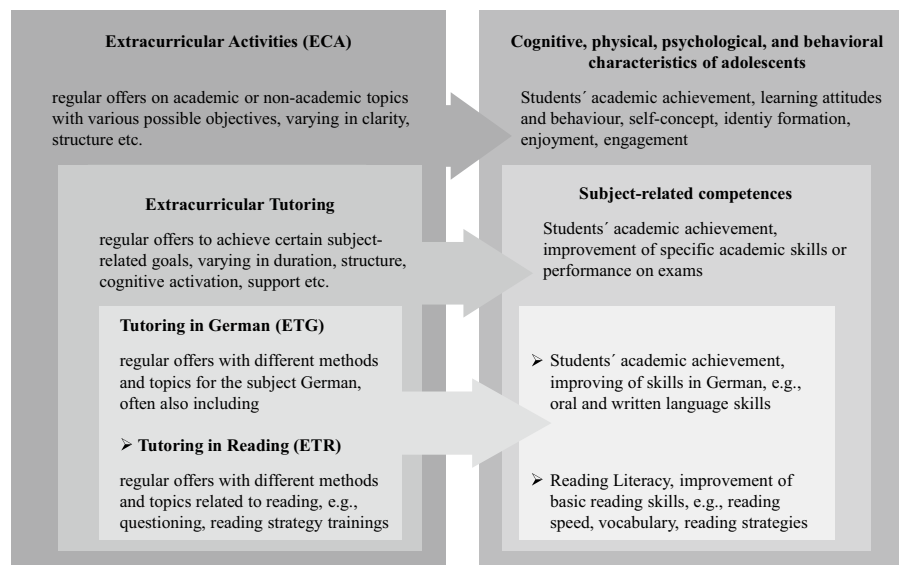
<sup>2</sup> Leibniz Institute for Educational Trajectories, Wilhelmsplatz 3, 96047 Bamberg, Germany

## Abstract

Because large-scale studies repeatedly indicated low reading literacy for many students, a need for interventions fostering reading literacy, such as extracurricular tutoring, has often been emphasized. Several reading promoting programs, suitable for extracurricular tutoring, were developed and shown to be effective in recent years. Moreover, these studies and analyses of extracurricular activities and tutoring yielded findings on learning-supporting characteristics and effects on students' achievement of such extracurricular offers. Nevertheless, little is known to date about the implementation of extracurricular tutoring in practice in Germany, particularly about its participation rates, characteristics, and effects on students' reading literacy. Thus, the present study investigated participation rates of extracurricular tutoring in reading and in German among students in general and among various subgroups between Grades 5 and 7. Furthermore, the study examined general and subject-specific characteristics and effects of extracurricular tutoring on reading literacy. The analyses used a sample of students ( $N = 5113$ ) of the *National Educational Panel Study* in Germany. In addition to descriptive analyses of characteristics and participation rates of extracurricular tutoring, effect sizes were calculated for comparing participation rates of subgroups of students. Furthermore, multi-group structural equation models were implemented to investigate average and differential effects of extracurricular tutoring, while controlling for initial reading literacy and other covariates. The results indicate that mainly students with low reading literacy and therefore a strong need for support participated in extracurricular tutoring, but the general participation rate was low. The descriptive outcomes showed a heterogeneous picture of tutoring offers and rare implementations of reading-promoting methods. Moreover, the results revealed no significant effects of extracurricular tutoring on students' reading literacy overall but suggested marginal differences for students with a low initial reading literacy. For future studies, more detailed assessments of characteristics and methods of extracurricular tutoring are recommended, particularly, in large-scale investigations on specific tutoring methods for students with support needs which may lead to improved guidelines regarding promising implementations of evidence-based reading promoting programs.

Reading literacy is central for individual and professional development throughout life, especially for learning, obtaining information, and enjoying texts in a variety of contexts. Referring to this fundamental importance, reading literacy is often defined as “understanding, using, reflecting on and engaging with written texts, in order to achieve one’s goals, to develop one’s knowledge and potential, and to participate in society” (OECD, 2009, p. 23). Despite its importance, many students gain only a low level of reading literacy during their school careers as shown repeatedly by various studies. For example, international large-scale assessments such as the *Progress in International Reading Literacy Study* (PIRLS) and the *Program for International Student Assessment* (PISA) as well as national assessments in Germany such as the *Internationale Grundschul-Lese-Untersuchung* (IGLU) indicated high rates of students with low reading literacy levels. With respect to the age cohort examined in the present study, in IGLU 2011 reading literacy of about 15.4% of Grade 4 students in Germany was below the basic level (level III; Tarelli et al., 2012). Looking at roughly the same cohort five years later, in PISA 2016 reading literacy of about 16% of Grade 9 students in Germany fell below the basic level (level II; Weis et al., 2016). Although age-group-specific differences in the criteria of reading literacy levels in IGLU and PISA have to be taken into account, both studies indicated a remarkable number of struggling readers in the respective age groups. Moreover, both studies showed that problems in reading occurred more often among boys than girls (Tarelli et al., 2012; Weis et al., 2016), although not significant in PISA 2015, and more often among students with foreign language backgrounds than native speakers (OECD, 2016; Tarelli et al., 2012). In addition to these cross-sectional studies, surveys with a longitudinal perspective provided results on the development of reading literacy across secondary school. For example, Pfost and Artelt (2013) revealed scissor effects in the development of students’ reading literacy between Grades 5 and 7 in academic and non-academic school types in Germany. During the observation period, the reading literacy levels of students drifted apart: while students with higher reading literacy in academic schools showed higher average growth in reading literacy, students in non-academic schools with lower reading abilities at the beginning achieved the lowest growth in their reading literacy (so-called “Matthew effect”).

However, to support students with low reading literacy and to prevent them from not being able to cope with later demands in school and beyond, all studies mentioned above emphasized the importance of promoting reading literacy, for example, by means of extracurricular tutoring (e.g., Artelt et al., 2010; Diedrich et al., 2019; Tarelli et al., 2012; Weis et al., 2016, 2019). To date, little is known about the extent to which students participate in appropriate extracurricular tutoring, the way such tutoring interventions are implemented on a broad scale and how effective they are. Therefore, the present study investigates the participation rates of students in general and with respect to specific subgroups. In addition, learning-supporting characteristics, and reading literacy-promoting effects of extracurricular tutoring are investigated with a longitudinal large-scale perspective.



**Fig. 1** Extracurricular offers with different characteristics and expected effects on students in previous research

## Theoretical background

### Concepts of extracurricular activities and tutoring

When reviewing previous research with a focus on participation rates, learning-supporting characteristics, and effects of extracurricular tutoring, one encounters a wide range of studies of diverse extracurricular offers that differ in their target criteria, underlying concepts, or terms. In particular, studies on *extracurricular activities*, *extracurricular tutoring* in various subjects, or *shadow education* respectively, and on evaluations of *reading promotion programs* most of which can be implemented as *extracurricular tutoring in reading* (see Fig. 1) provide insights relating to the assumptions about the research questions investigated in this study. While all of these types of extracurricular offers are regularly attended by students outside of classes for a period of time, for example, they may have different goals (e.g., Baker et al., 2001).

While *extracurricular activities* (ECA) address academic subjects or non-academic support and enrichment, such as sports and arts (e.g., Feldman & Matjasko, 2005; Shulruf & Wang, 2013), *extracurricular tutoring* is mainly characterized as subject-related and often as (parent-) paid support (e.g., Bray, 2014; Guill et al., 2020b; Mischo & Haag, 2002). This is especially true in Asian cultures, where students often receive extracurricular tutoring—also referred to as *shadow education*—for a fee to improve their academic achievement (Bray, 2014). In contrast, in European cultures, and Germany in particular, there are many free offers of extracurricular tutoring in addition to paid tutoring offers, both of which aim to promote academic achievement in specific subjects. For example, *extracurricular tutoring in reading* aims to promote students' reading literacy, for which *reading promotion programs* are particularly suitable, many of which have been developed and successfully evaluated in individual studies or meta-analyses in recent years (e.g., Almasi & Palmer, 2013; Baye et al., 2018; Holopainen et al., 2018; Souvignier & Antoniou, 2007; Suggate, 2016).

Given the heterogeneity of extracurricular offers, *extracurricular tutoring* in this study refers to all interventions—regardless of whether they are free or fee-based and offered by schools or non-school institutions—in which students participate regularly and for an agreed period of time outside of regular classes in order to improve their academic performance in specific subjects. Concerning such extracurricular tutoring to improve students' achievement in German (ETG) and reading (ETR), this study investigates (a) students' participation rates, (b) learning-promoting characteristics, and (c) effects on students' reading literacy. With focus on these three aspects, our research assumptions are based on previous results from the studies on various extracurricular offers which are mentioned above and outlined in the following three sections as basis of our research questions.

### **Participation rates of extracurricular activities and tutoring**

To date, only a few international studies provided results on rates of students' participation in *extracurricular offers* worldwide. For example, the 1994/95 *Trends in International Mathematics and Science Study* (TIMSS) found that about 20% of students in Grades 7 and 8 in the involved countries participated in extracurricular tutoring in mathematics which was offered in a wide variety of forms (e.g., Baker et al., 2001). Regarding participation rates in Germany, supplemental studies in TIMSS and IGLU based on parent surveys indicated that rates of students participating in extracurricular tutoring in Grade 4 declined between 2006 and 2015 (Guill & Wendt, 2016): participation rates in extracurricular tutoring in German decreased from about 22% (2006) to 19% (2011) and finally to 15% (2015). Concerning extracurricular tutoring in reading and spelling, participation rates dropped from about 19% (2006) to 16% (2011) and finally to 14% (2015). The assumption that the decrease in participation rates of tutoring observed can be attributed to the simultaneous increase in the number of offers and their use in all-day schools was not confirmed in subsequent analyses (Guill & Wendt, 2016): accordingly, students in all-day schools with corresponding support programs did not differ from students in regular schools without all-day programs in terms of their participation rates in tutoring. Thus, the development of the participation rates of tutoring was unrelated to the offer and its utilisation in all-day schools; students in all-day schools showed no lower participation rates in extracurricular tutoring offers (Guill & Wendt, 2016). Concerning students of Grade 5 in academic school types in Germany, the longitudinal project *Ganz In* in North Rhine-Westphalia (Germany) revealed that about 7% of them attended extracurricular tutoring in German (Guill et al., 2020b). Among students between Grades 7 and 8 in Hamburg (Germany), the *Study of Student Competencies and Attitudes* (KESS) found that about 19% of them, including mostly boys, had recently participated in extracurricular tutoring, mostly lasting about 6 months and lasting a maximum of 2 h per week (Guill et al., 2022). Concerning participation rates in Grade 9, Hawrot's study (2024) based on data from the *National Educational Panel Study* (NEPS; Starting Cohort 4 in 2010) indicated that about 3% of students attended paid tutoring in German.

Regarding subgroups of students and their reasons for participation in tutoring, according to the TIMSS in Germany, the majority of students participated in tutoring in mathematics because of low achievement, and, thus, for remedial purposes (Baker et al.,

2001). In addition, the likelihood of participation increased when grades were weaker. Concerning the characteristics of participants of paid tutoring, Hawrot (2024) found in her study on students in Grade 10 different individual characteristics—specifically, their helplessness, subjective task values, past school achievement, mother tongue, and school type—predicting their participation in tutoring in math and German. In particular, students' prior school achievement—operationalized as reversed grades in German and mathematics in Grade 10—was associated with their participation in paid tutoring in German, which therefore was used also primarily for remedial purposes: students with better grades in German showed a lower likelihood to participate in paid tutoring in German. Moreover, increasing achievement in German at the level of one grade in the mid-term certificate was associated with a decreasing likelihood of participation in paid tutoring in German by 40%. Similar associations between cultural capital at home as well as grades and participation in paid tutoring were found for students in Grade 5 (Guill et al., 2020b). Additionally, students in academic tracks had a lower likelihood of participating in paid tutoring in German by 55%, compared to students of non-academic school tracks (Hawrot, 2024). Furthermore, in this study, a foreign language background and a higher socio-economic background were associated with higher likelihoods of participation in paid tutoring in German, whereas gender and domain unspecific cognitive ability did not matter.

#### **Characteristics of extracurricular activities and tutoring**

Concerning learning-supporting characteristics of *extracurricular activities*, Hattie (2009) concluded that the strongest effects on academic achievement are produced by pursuits that induce academic activities related to the objects of interest. According to his assumptions, students engage voluntarily, joyfully and persistently with respective topics, learn about them with and from others and are also strengthened in terms of performance and motivation, which in turn is associated with enhanced achievements. Moreover, according to Lewis (2004), extracurricular activities with the highest effects are characterized by clear structures, organization, and regularity.

Concerning learning-supporting characteristics of *subject-related extracurricular tutoring*, previous research revealed different results. For example, results from TIMSS in 1994/95 showed that tutoring in mathematics in most participating countries had a duration of about 2 h per week or less (Baker et al., 2001). Moreover, Guill et al. (2020a)—examining structure, cognitive activation, and support—found no effects of the instructional quality of extracurricular tutoring in German, Mathematics or other subjects on students' grades in mathematics or German.

With respect to *reading promoting programs* suitable for implementation in *extracurricular tutoring in reading* (ETR), Almasi and Palmer (2013) reported the largest effects for small interactive groups with loud thinking and strategy modelling by tutors. Similar to Souvignier and Antoniou (2007, see below), they found the largest effects on study-specific measures designed by the researchers of the study. In line with the results reported above, cooperative learning as well as one-to-one instruction—beyond small group and whole-school interventions—showed the highest effects in a meta-analysis of experimental evaluated reading-promoting programs in secondary school (Baye et al., 2018). Moreover, this study revealed that students benefitted mostly from social

and cognitively activating as well as motivating programs, but less of additional reading time or technology-based reading instruction. Concerning social settings of reading promoting programs, the study of Souvignier and Antoniou (2007) distinguished between interventions implemented in classrooms versus in special pedagogical contexts and outside of regular classes respectively. Their results referring to standardized tests or study-specific tests showed that the latter was more effective. Concerning results from both test types, they found larger effects for interventions with a duration of more than 12 h, compared to shorter interventions, and no effects due to types of tutors, distinguishing between teachers, special education teachers, and researchers. In contrast to these results on types of tutors, Suggate (2016) reported the largest effects in reading promoting programs for trained tutors, followed by peers and finally teachers as compared to instruction by preschool teachers, technology applications, or researchers. Furthermore, in this study, individual tutoring in reading outperformed tutoring in groups in terms of their effects in post- as well as in follow-up-tests. In contrast, Holopainen et al. (2018) found that faster development of reading-related skills occurred with small-group interventions that lasted about 38 h per year, but not longer, compared to individual interventions with a higher amount of time (more than 48 h per year). In their study with Finnish students with learning difficulties between Grades 1 and 2, they investigated the effects of part-time special education methods implemented by several types of tutors, in various social settings, and with different amounts with respect to competences in reading and spelling.

#### **Effects of extracurricular activities and tutoring**

In general, it is assumed that participation in extracurricular activities and tutoring leads to improvements in the corresponding performance areas. Concerning *extracurricular activities*, several studies indicated strong associations between students' participation and their educational achievement, beyond effects on dropout rates from school and enrolment in tertiary education (e.g., Feldman & Matjasko, 2005; Shulruf & Wang, 2013). For example, Lewis' (2004) meta-analysis distinguishing between six types of activities revealed the highest positive effects on academic achievement—in addition to positive effects on engagement, identity formation, and other aspects—from participation in general extracurricular activities compared to sports, arts, working and vocational, pro-social and community-related activities. However, this study did not provide more detailed information on the individual methods and topics of the 'general extracurricular activities' studied. Moreover, longitudinal cross-lagged-panel models from Grade 3 until Grade 8 (Carbonaro & Maloney, 2019) indicated very small—but with ascending grades increasing—positive effects of participation in extracurricular activities of various types on later academic abilities. According to Shulruf's (2010) meta-analysis, most studies in the United States showed positive associations between students' participation in extracurricular activities and academic achievement, but evidence of causal effects is still lacking. Accordingly, Shulruf and Wang (2013) stated that to date results on participation in extracurricular activities provided less evidence of expected effects on students' achievement. They therefore emphasized the urgent need for further research, particularly for investigating causal characteristics and effects of extracurricular activities. Furthermore, Shulruf (2010) mentioned methodological issues

of previous studies—in particular, various underlying approaches, outcome definitions, and criteria—that need to be optimized to detect supporting characteristics and causal effects of extracurricular activities.

Concerning the effectiveness of *subject-related extracurricular tutoring*, previous studies showed an inconsistent picture. For example, Mischo and Haags' (2002) quasi-experimental study showed that grades in mathematics, English, Latin, or French improved significantly when students of different ages (Grade 5 to Grade 11) participated in paid tutoring provided by institutions four days a week for about 90 min over a nine-month period. In contrast, the longitudinal study of Guill et al. (2020a) on students in Grade 10 indicated no positive effects of paid tutoring in German, mathematics or other subjects on students' grades in respective subjects. Instead, regression of German grades at the end of the year on participation in tutoring revealed a low negative effect, also when various variables (i.e., Midterm grade, general cognitive ability, perceived helplessness, interests, academic aspirations [seeking for general qualification for university entrance], gender, HISEL, academic school type, and prior tutoring) were controlled for. Overall, all of these control variables—except from academic aspirations—significantly predicted academic achievement, e.g., with highest positive strong effects of Midterm grades and highest negative low effects of helplessness. However, gender and school types played a marginal role for Germans grades of tutoring participants, with better grades of girls compared to boys and also for students in non-academic schools compared to academic school students. Accordingly, the KESS study showed no positive effects of extracurricular tutoring in German, Mathematics, English and learning strategies—neither as a function of duration (month), intensity (weekly scope), instructional focus, nor due to students' motivation—on subject-specific grades or test achievement, even after controlling relevant covariates (Guill et al., 2022). Moreover, some specific features of extracurricular tutoring in German, e.g., tutoring intensity, revealed low negative effects on aspects of academic achievement. Therefore, the authors of both studies concluded that paid tutoring provides no effective strategy to improve academic achievement. Accordingly, in about 40 of the involved countries in TIMSS, no effects were found from participation in extracurricular tutoring in mathematics on students' math achievement (Baker et al., 2001).

With respect to *reading promoting programs*, many studies revealed positive results on reading achievement (e.g., critical review on several reading promoting programs by Almasi & Palmer, 2013; Baye et al., 2018; Holopainen et al., 2018; Souvignier & Antonio, 2007; Suggate, 2016). For example, the meta-analysis by Edmonds et al. (2009) showed large effect sizes for proficient, struggling, and disabled readers in Grade 6 to Grade 12 for multicomponent interventions and even larger effects for comprehension interventions. These effective comprehension interventions included explicit instruction, teaching how to use multiple strategies with authentic texts, how to monitor, self-question, and regulate comprehension as well as opportunities to exercise the use of strategies. Similarly, the meta-analysis by Souvignier and Antoniou (2007) revealed the largest positive effects on reading literacy of students with reading disabilities or learning problems—measured by means of standardized versus study-specific tests—for interventions on question strategies. In this study, the second most effective programs gained to improve reading skills or complex strategies, while text-enhancing programs

were less effective. In line with this, also the meta-analysis on long-term and follow-up effects of reading interventions by Suggate (2016), revealed the most lasting and transferring effects for comprehension and phonemic awareness interventions.

### Research questions

Against this background, the first research question (RQ) addresses students' general participation rates for extracurricular tutoring in German (ETG) and reading (ETR) between Grades 5 and 7 (RQ1.1). Referring to the results of IGLU and TIMSS presented above—in particular, for 2010 when the students that are the focus of the study were in Grade 5—we expect a participation rate of about 19% of all students for ETG (Guill & Wendt, 2016), and a lower rate of about 16% for ETR (Guill & Wendt, 2016). In addition, we investigate participation rates in ETG and ETR of students from subgroups (RQ1.2) that differ in terms of migration background, gender, cultural capital at home, German grades in Grade 5, school types, and prior reading literacy, including reading difficulties. Referring to the results of IGLU (Tarelli et al., 2012) and PISA (Weis et al., 2016), reading literacy of about 15% of students in Grade 4 and 16% in Grade 9 fell below the basic level and indicated a strong need for support (Artelt et al., 2010; Diedrich et al., 2019; Weis et al., 2019). Therefore, we assume these students (about 15–16% of all students), especially boys and students with a migration background, to make up the main target group of extracurricular tutoring in German and reading, and ideally participate in ETG or ETR, respectively. This expectation is partly supported by previous results showing higher probabilities of participation in paid tutoring in German for students with migration background (Guill et al., 2020b; Hawrot, 2024). However, results on gender-specific participation rates are inconsistent: while some studies show that most participants are boys (Guill et al., 2022), in others gender is unrelated to participation in paid tutoring in German (Hawrot, 2024). Further results support our expectation that mainly students with poor grades in German participate in ETG (Guill et al., 2020b; Hawrot, 2024). Concerning students in academic school types, we expect a participation rate in ETG of about 7% for students of Grade 5 (Guill et al., 2020b), and a smaller number of participants in higher Grades, referring to lower rates shown in previous studies for students in Grade 9 (about 3%; Hawrot, 2024). In comparison to the participation of students in academic schools, we expect higher participation rates in extracurricular tutoring for students of non-academic schools (Hawrot, 2024).

In our second research question, we investigate learning-supporting characteristics of extracurricular tutoring (in German) from a descriptive view (RQ2). In particular, we describe social settings, types of tutors, locations, and weekly amounts. Referring to previous findings, we assume that individual (Baye et al., 2018; Suggate, 2016) and small group tutoring in reading (Almasi & Palmer, 2013; Holopainen et al., 2018) are the most effective social forms to promote reading and should therefore be implemented. In addition, findings from previous studies suggest that trained tutors, peers or teachers, are the most effective tutors compared to others (Suggate, 2016), and, thus, observing these types of tutors in the data would be desirable. The weekly amount and intensity of tutoring in various subjects for which effects on grades have been shown was about 6 h per week over a nine months period (Mischo & Haag, 2002) which would therefore be desirable. However, the intensity and weekly amount of effective tutoring in reading



in reading promoting programs lasted at least 12 h (Souvignier & Antoniou, 2007) and did not exceed 38 h per year, that is, took place about once a week over a school year (Holopainen et al., 2018).

In our third research question, we analyze the effects of students' participation in ETG on their reading literacy (RQ3.1). With reference to previous results on the effects of extracurricular activities (e.g., Carbonaro & Maloney, 2019; Feldman & Matjasko, 2005; Lewis, 2004), paid tutoring (Mischo & Haag, 2002), and, in particular, reading promotion programs (e.g., Almasi & Palmer, 2013; Baye et al., 2018; Edmonds et al., 2009; Holopainen et al., 2018; Souvignier & Antoniou, 2007; Suggate, 2016), most of which can be implemented as ETR, we expect positive effects of participation on students' reading literacy. In contrast, other results do not support such expectations (Baker et al., 2001; Guill et al., 2022; Shulruf, 2010; Shulruf & Wang, 2013), particularly not for effects of extracurricular tutoring in German on grades (Guill et al., 2020a, 2022). We also investigate the differential effects of ETG on the reading literacy of students of subgroups that differ with respect to various sociodemographic characteristics, academic achievement, and prior reading literacy (RQ3.2). Specifically, we examine the extent to which students with high support needs—particularly struggling readers, boys, and students with migration backgrounds—benefit from ETG according to reported recommendations (Artelt et al., 2010; Diedrich et al., 2019; Tarelli et al., 2012; Weis et al., 2016, 2019). Referring to previous results on the effects of reading promoting programs (Almasi & Palmer, 2013; Baye et al., 2018; Edmonds et al., 2009; Edmonds et al., 2009; Holopainen et al., 2018; Souvignier & Antoniou, 2007; Suggate, 2016), we expect compensatory effects of ETG, particularly for struggling readers (e.g., Edmonds et al., 2009; Souvignier & Antoniou, 2007).

## Method

### Sample and procedure

The present study is based on the longitudinal, multi-cohort *National Educational Panel Study* in Germany (NEPS; Blossfeld & Roßbach, 2019) that follows different representative samples across their life courses. The present study focuses on a sample of students initially attending Grade 5 in secondary schools that were subsequently examined in Grades 6 and 7 (SC3: starting cohort 3). Most of these students were tested in small groups at their respective schools by trained test administrators (for further information on data collection and test administration see NEPS-Network, 2021; Steinhauer et al., 2015). In total,  $N=7552$  students participated in one of the reading literacy assessments (either in Grade 5 or 7). Out of this sample, all students were included in our study that provided information on their participation in a tutoring program ( $N=5113$ , 67.70% of the students who provided competence data).<sup>1</sup> This resulted in 437 students who had participated in ETG during the study period of 3 years and a comparison group of 4676 students who did not participate in ETG (i.e., no participation at all or participation in

<sup>1</sup> In contrast to the complete NEPS-SC3 data, the selected sample is not representative for all German students in Grade 5 to 7 at the assessment time. Information on the participation in extracurricular tutoring is essential for our study and we excluded participants who did not provide any information on this, because a replacement of the missing data is not meaningful (e.g., Gomila & Clark, 2022, who recommend model-based strategies for handling missing data only for the outcome variable or for covariates, but not for the treatment indicator).

tutoring in another subject). The students included 2454 (48%) girls. About 21% of them had a migration background. Their cultural capital (Sieben & Lechner, 2019) and home literacy environment respectively, as an important factor in literacy acquisition (Buhl & Hilkenmeier, 2016; McElvany et al., 2009), reflected in the number of books at home, was rather high (60% reported to own more than 100 books at home). Almost half of the sample (49%) attended academic school types (i.e., higher secondary education, e.g., gymnasium).

### Instruments

Reading literacy was measured in Grades 5 and 7 with standardized achievement tests that were specifically constructed for administration in the NEPS. These instruments included several texts with closed-response items that, followed the theoretical framework by Gehrler et al. (2013), and referred to three cognitive requirements (i.e., identification of single facts, inference of conclusions, evaluation, and interpretation). The tests included 32 and 40 dichotomous and polytomous items in Grades 5 and 7, respectively. Both tests were administered on paper with a time limit of 28 min, scaled using a one-parametric item response model (Masters, 1982), and linked across points of measurement to allow for meaningful longitudinal comparisons (Fischer et al., 2016). Detailed results on the psychometric properties of the tests in the present sample including item fit and tests for differential item functioning are documented in technical reports (Krannich et al., 2017; Scharl et al., 2017). The marginal reliabilities at the two measurement points were 0.78 and 0.79, respectively. Point estimates of students' competences were derived as weighted likelihood estimates (WLE; Warm, 1989) which are also available in the scientific use files (NEPS Network, 2021a). We use the WLE scores as a proxy for each student's reading literacy in the descriptive analysis. However, as the tests are not perfectly reliable, the WLE scores include error components that can bias effect estimates (see e.g., Lechner et al., 2021; Sengewald & Mayer, 2024). The NEPS data include the item responses for each achievement test, too, and we used the item-level data to correct for measurement error in reading literacy, in the form of plausible values (PV; Mislevy, 1991). As such, we specified a two-dimensional partial credit model (Masters, 1982) and drew 20 PVs at each measurement point for each student that are used in the effect analysis (see e.g., Sengewald & Mayer, 2024).

Information on students' participation in extracurricular tutoring in German and reading was derived from the parent interviews (see interview questions 1, 2 and 3 in Table 5 in Appendix). The respective binary index reflected whether a student participated in ETG in Grades 5, 6, or 7 or not (0 = no participation, 1 = participation). Furthermore, information was available on several characteristics of tutoring taken in Grade 6, that is, locations, social settings, and types of tutors.

Finally, various background information on the students assessed in interviews of target persons or parents was considered. This included students' gender (0 = male, 1 = female), migration background (0 = no, 1 = yes; see Table 6 in Appendix), cultural capital (0 = 100 books or less at home, 1 = more than 100 books at home), school type (1 = gymnasium, 0 = other), and their school grades in German in Grade 5 (on a scale from 1 = best to 6 = failing).

### Handling of missing values

We constructed our sample so that all students provided information on their participation in a tutoring program in Grades 5, 6, or 7. However, further information on the characteristics of tutoring was only assessed in Grade 6 and thus was not available for all students. We report missing values in tutoring characteristics and discuss the restrictions of the database.

Furthermore, we consider only students, who participated in at least one of the reading literacy assessments, either in Grade 5 or Grade 7. We use a model-based strategy for handling missing competence data (i.e., PV estimation in a two-dimensional model). As such, the model includes the reading literacy of the other measurement occasion and all additional covariates as background information for replacing the missing values in reading literacy.

Because the considered covariates used in the causal effect analyses were rather time-stable characteristics, we replaced missing values by the available information of subsequent waves (i.e., we carried the last observation forward to impute missing values of migration background, gender, cultural capital captured as the number of books at home, and school type). For school grades in Grade 5, we imputed missing values in students' self-reports with the respective information from the parent interview. The remaining missing values in the covariates (i.e., migration background, cultural capital, school type, and grades that are summarized in Table 1), were imputed using classification and regression trees (Breiman et al., 2017) to provide complete background data for the effect analysis. Because the number of missing values was small for most of the covariates, we used only a single imputation round to replace the missing covariate values.

### Statistical analyses

#### *Comparative description of participation rates and characteristics of tutoring in German and reading*

To investigate questions on participation rates (RQ1.1) and learning-supporting characteristics of tutoring in German and reading (RQ2), we conducted descriptive analyses. In addition, participants of ETG and ETR compared to students without participation were described in terms of participation rates of students of individual subgroups concerning migration background, gender, cultural capital, school type, German grades, and reading literacy in Grade 5 (RQ1.2). Moreover, we estimated differences between participating rates of ETG and ETR of these subgroups in terms of effect sizes: for categorical variables (i.e., migration background, gender, cultural capital, and school type) we calculated odds ratios and for metric variables (i.e., for grades in German, initial reading literacy) Cohen's  $d$ , respectively. These analyses were based on WLEs as proficiency scores of reading literacy.

#### *Analyses of treatment effects*

For the effect analyses (RQ3.1), reading literacy in Grades 5 and 7 was modeled in the form of PVs to correct for measurement error (see e.g., Lechner et al., 2021; Senge-wald & Mayer, 2024). The PVs for Grade 7 served as the outcome, in the comparison

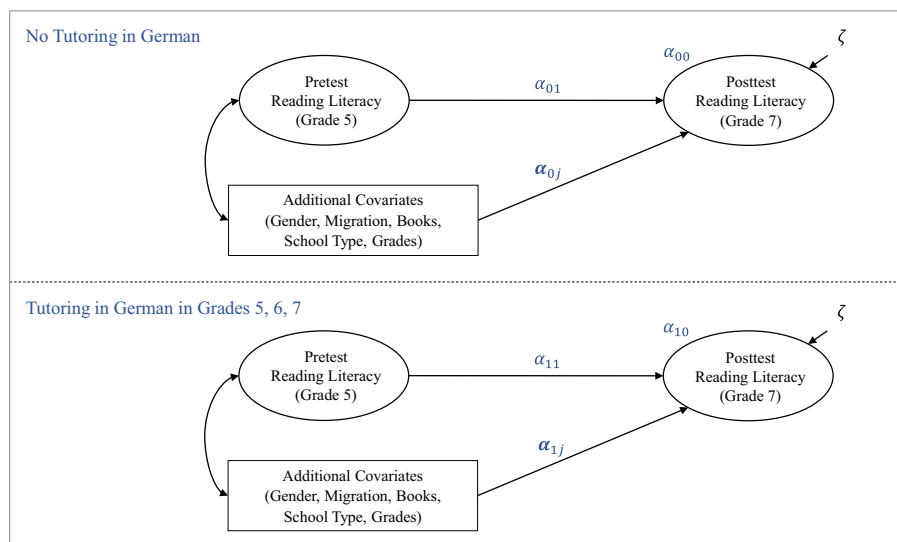
**Table 1** Participation rates of extracurricular tutoring in German of students of various subgroups

| Background variables      | Total | Missings   | Extracurricular tutoring in German |           |                       | Extracurricular tutoring in reading |          |                       |
|---------------------------|-------|------------|------------------------------------|-----------|-----------------------|-------------------------------------|----------|-----------------------|
|                           |       |            | No                                 | Yes       | Effect size [95% CI]  | No                                  | Yes      | Effect size [95% CI]  |
| Migration background      | 5040  | 73 (1%)    |                                    |           | 2.16* [1.74, 2.67]    |                                     |          | 2.31* [1.69, 3.13]    |
| Migration                 | 1066  |            | 917 (86%)                          | 149 (14%) |                       | 998 (94%)                           | 114 (6%) |                       |
| German                    | 3974  |            | 3696 (93%)                         | 278 (7%)  |                       | 3860 (97%)                          | 68 (3%)  |                       |
| Gender                    | 5113  | 0 (0%)     |                                    |           | 0.61* [0.50, 0.75]    |                                     |          | 0.64* [0.47, 0.86]    |
| Boys                      | 2659  |            | 2384 (90%)                         | 275 (10%) |                       | 2541 (96%)                          | 118 (4%) |                       |
| Girls                     | 2454  |            | 2292 (93%)                         | 162 (7%)  |                       | 2383 (97%)                          | 71 (3%)  |                       |
| Cultural capital          | 5062  | 51 (1%)    |                                    |           | 0.67* [0.55, 0.82]    |                                     |          | 0.51* [0.38, 0.69]    |
| Low ( $\leq 100$ books)   | 1999  |            | 1792 (90%)                         | 207 (10%) |                       | 1897 (95%)                          | 102 (5%) |                       |
| High ( $> 100$ books)     | 3063  |            | 2842 (93%)                         | 221 (7%)  |                       | 2981 (97%)                          | 82 (3%)  |                       |
| School type               | 5105  | 8 (0%)     |                                    |           | 0.38* [0.30, 0.47]    |                                     |          | 0.26* [0.18, 0.36]    |
| Academic                  | 2508  |            | 2385 (95%)                         | 123 (5%)  |                       | 2469 (98%)                          | 39 (2%)  |                       |
| Non-academic              | 2597  |            | 2284 (88%)                         | 313 (12%) |                       | 2447 (94%)                          | 150 (6%) |                       |
| German grades in Grade 5  | 3918  | 1195 (23%) |                                    |           | 0.82* [0.72, 0.93]    |                                     |          | 0.92 [0.75, 1.08]     |
| Grade 1 (best)            | 565   |            | 559 (99%)                          | 6 (1%)    |                       | 564 (100%)                          | 1 (0%)   |                       |
| Grade 2                   | 1789  |            | 1691 (95%)                         | 98 (5%)   |                       | 1756 (98%)                          | 33 (2%)  |                       |
| Grade 3                   | 1232  |            | 1046 (85%)                         | 186 (15%) |                       | 1156 (94%)                          | 76 (6%)  |                       |
| Grade 4                   | 308   |            | 231 (75%)                          | 77 (25%)  |                       | 267 (87%)                           | 41 (13%) |                       |
| Grade 5                   | 22    |            | 15 (68%)                           | 7 (32%)   |                       | 19 (86%)                            | 3 (14%)  |                       |
| Grade 6 (failing)         | 2     |            | 1 (50%)                            | 1 (50%)   |                       | 2 (100%)                            | 0 (0%)   |                       |
| Reading literacy (WLE)    | 3833  | 1280 (25%) |                                    |           | -0.62* [-0.73, -0.51] |                                     |          | -0.75* [-0.92, -0.58] |
| -2SD (struggling readers) | 62    |            | 52 (84%)                           | 10 (16%)  |                       | 55 (89%)                            | 7 (11%)  |                       |
| -1SD                      | 543   |            | 434 (80%)                          | 109 (20%) |                       | 491 (90%)                           | 52 (10%) |                       |
| +1SD                      | 645   |            | 627 (97%)                          | 18 (3%)   |                       | 641 (99%)                           | 4 (1%)   |                       |
| +2SD (top readers)        | 122   |            | 119 (98%)                          | 3 (2%)    |                       | 122 (100%)                          | 0 (0%)   |                       |

*N* = 5113. Effect sizes for comparisons of different treatment groups on the background variables represent odd ratios (for migration background, gender, cultural capital, and school type) and Cohen's *d* respectively (for German grades in Grade 5, reading literacy). For reading literacy, weighted likelihood estimations (WLE; without imputation of missings) from scientific use file (NEPS Network, 2021a) were used

\**p* < 0.05

of the treatment groups. Specifically, we were interested in the difference in reading literacy between students participating in ETG (i.e., the treatment group) and students without participation in ETG (i.e., the control group). However, baseline differences between the groups can bias effect estimates in this non-randomized group comparison. We used a multi-group structural equation modeling framework—the *EffectLiteR* framework (Mayer et al., 2016 with extensions from Sengewald & Mayer,



**Fig. 2** Estimating the adjusted effects of tutoring in German on reading literacy in Grade 7. *EffectLiteR* estimates a multigroup structural equation model that separates the two tutoring groups and estimates group specific regression coefficients for predicting the latent posttest reading literacy (i.e., plausible values from the reading literacy test in Grade 7) given all covariates. The set of covariates contains the latent pretest reading literacy (i.e., plausible values from the reading literacy test in Grade 5) and the vector of five additional covariates with  $j = 2, \dots, 6$  respectively

2024)—to account for baseline group differences via covariate adjustment (see Fig. 2). For covariate adjustment, we used different sets of covariates, that is, either (a) no covariates, (b) only the pretest (i.e., the PVs for the initial reading literacy in Grade 5), or (c) the pretest and additional covariates (i.e., migration, gender, cultural capital, school type, grades) to investigate the relevance of the covariates. In the effect analyses, we then controlled for all relevant covariates.

To examine the differential effects of tutoring, we estimated different types of effects following Mayer et al. (2020): first, we considered the average treatment effect (AE) that describes the (covariate-adjusted) difference between the treatment groups for all students under investigation. In addition, we estimated various conditional treatment effects (CE), which describe the (covariate-adjusted) difference between the treatment groups for different subgroups. In particular, we examined differences in the treatment effect by gender (i.e., for boys and girls), migration status (i.e., for students with or without migration background), and different pretest values (i.e., for reading literacy in Grade 5 one or two standard deviations below  $[-1SD, -2SD]$  or above  $[+1SD, +2SD]$  the mean). Furthermore, we explored conditional effects by treatment status and report the effects on the treated (i.e., for those participating in ETG) and on the non-treated (i.e., for the control group). Finally, in a third step, we also estimated conditional effects for the combination of the treatment status with the other covariates. This can be particularly informative because the subgroups of treated and non-treated students integrate the specific characteristics of students who do not or do participate in ETG. Further differentiating their gender, migration background, and pretest abilities allowed for more detailed insights into for whom ETG is most effective.

### **Software and open material**

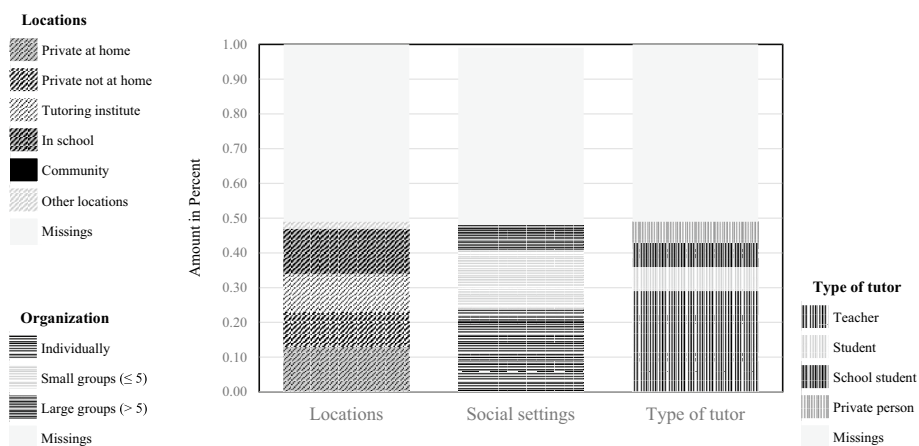
The measurement model for reading literacy and the plausible values were estimated in *R* (Version 4.2.1; R Core Team, 2022) with *TAM* (Version 3.7–16; Robitzsch et al., 2022). Missing values were imputed with *mice* (Version 3.14.0; van Buuren et al., 2011). The treatment effects were obtained with *EffectLiteR* (Version 0.4–5.015; Mayer et al., 2016) which relied on *lavaan* (Version 0.6–9; Rosseel, 2012). The raw data is available after registration at NEPS Network (2021a), whereas the analysis code that allows for reproducing our findings is provided at [https://osf.io/2su98/?view\\_only=8e44cdd35ee047d59e120d03621e4ff2](https://osf.io/2su98/?view_only=8e44cdd35ee047d59e120d03621e4ff2).

## **Results**

### **Participation rates and characteristics of extracurricular tutoring in German and reading**

Concerning general participation rates (RQ1.1), during the 3-year study period, 437 students (i.e., about 9% of our total sample) participated in ETG and 4676 students (i.e., about 92% of our total sample) did not participate. Concerning the content of this tutoring, 186 students (i.e., about 43% of the group with ETG; and about 4% of our total sample) participated in reading or understanding exercises, and thus in ETR. Furthermore, results show that 236 students (i.e., about 54% of the group with ETG) did not participate in ETR, and 12 persons reported no information on ETR (i.e., 0.03% of the group with ETG had a missing value). Among students participating in ETG, most parents reported short-term participation. Specifically, about 76% of the group with ETG participated during one wave of the survey ( $N=333$ ), 19% participated during two waves ( $N=85$ ), and for about 4% of students ( $N=19$ ) extracurricular tutoring in German during all three assessment waves was reported.

Participation rates of students of different subgroups in our sample ( $N=5113$ ) who provided information on their participation in extracurricular tutoring in German and reading are presented in Table 1 (RQ1.2). First, the results show that participation rates in the different subgroups differ significantly in most comparisons (see columns of Effects Sizes): accordingly, students with a migration background were significantly more likely to participate in ETG and also ETR than students with a German background. Specifically, about 7% of students with a German background took part in ETG and 3% in ETR, whereas 14% of students with a migration background took part in ETG and 6% in ETR. Similarly, the likelihood of participating in ETG as well as in ETR was significantly higher for boys (10%; 4%) than for girls (7%; 3%) and also for students with low cultural capital (10%; 5%) than for students with high cultural capital (7%; 3%). Furthermore, results show that rates of participation in ETG and ETR in academic school types were significantly lower than in non-academic school types: while about 5% of students of academic schools participated in ETG and 2% in ETR, about 12% of students of non-academic schools received ETG and 6% ETR. As far as grades in German in Grade 5 are concerned, the majority of the participating students had a grade of 3 or better. However, the likelihood of participating in ETG but not in ETR was significantly related to grades in German. Specifically, of the students whose grade in German was weaker ( $N=332$ ), about 26% ( $N=85$ ) participated in ETG and about 13% ( $N=44$ ) participated in ETR. Finally,



**Fig. 3** Characteristics of extracurricular tutoring in German and other subjects

the likelihood of participating in ETG and ETR was significantly related to levels of reading literacy. Based on the WLEs, poor readers (with reading literacy one or two standard deviations below the mean) made up about 16% of the sample. Among students with very low reading literacy and struggling readers respectively ( $-2SD$ :  $N=62$ ), about 16% participated in ETG and 11% in ETR. Concerning students with low reading literacy ( $-1SD$ :  $N=543$ ), about 20% received ETG and 10% participated in ETR. In contrast, about 3% of students with high reading literacy ( $+1SD$ :  $N=645$ ) took part in ETG and 1% of them in ETR. Concerning students with very high reading literacy ( $+2SD$ :  $N=122$ ), about 2% participated in ETG and none in ETR.

Concerning characteristics of extracurricular tutoring in German and other subjects (RQ2), results are provided in Fig. 3. Since information on the tutoring setting was only assessed in the second assessment wave, descriptive information is only available for tutoring in Grade 6. Because the parent-reported information contains many missing values (51%), the respective results refer to the selected sample of students receiving ETG for which their parents reported information on ETG: most frequently reported locations of tutoring (13%) were private homes and places at school, followed by institutes (11%) or other nonformal places outside from home (10%). In contrast, tutoring was rarely implemented in other private settings (2%) or in other institutions ( $N=2$ ). Concerning social settings of tutoring, the most common was individual tutoring (24%), followed by tutoring in small groups (17%). Furthermore, students rarely participated in tutoring with large groups (7%). Regarding the type of tutors, typically the tutors were teachers (29%), but rarely students or school students (7%) or further private persons (6%). In terms of the number of tutoring hours per week for tutoring in all subjects, parents of about 75% of students with ETG reported that their children spent one or two hours per week for all tutoring offers ( $N_{1h}=185$ , i.e., 42% of the sample with ETG;  $N_{2h}=144$ , that is, 33% of the sample with ETG). Three hours per week were reported for 11% of the students with ETG ( $N_{3h}=48$ ) and four hours per week for 7% of the same group of students ( $N_{4h}=29$ ). Finally, 6% of students with ETG participated 5 h or more per week in tutoring ( $N_{5h+}=27$ ).

**Table 2** Average treatment effects of tutoring in German for different sets of covariates

| Covariates                   | Est.   | SE    | Est./SE | p-value | Effect size |
|------------------------------|--------|-------|---------|---------|-------------|
| None                         | -0.568 | 0.050 | -11.356 | <0.001  | -0.606      |
| Pretest                      | -0.067 | 0.068 | -9.991  | 0.328   | -0.072      |
| Pretest and other covariates | -0.017 | 0.068 | -0.246  | 0.807   | -0.018      |

Pretest = reading literacy in Grade 5 (PV); other covariates = migration background, gender, cultural capital, school type, and grades in German in Grade 5. We report unstandardized effect estimates (Est.), robust standard errors (SE), tests of estimates against zero (Est./SE and p-values), and effects in standard deviation units of the outcome

**Table 3** Conditional treatment effects of tutoring in German

| Grouping         | Effect                   | Est    | SE    | Est./SE | p-value | Effect size | Δ effect size |
|------------------|--------------------------|--------|-------|---------|---------|-------------|---------------|
| Treatment status | CE <sub>x=control</sub>  | -0.021 | 0.072 | -0.298  | 0.767   | -0.023      | 0.059         |
|                  | CE <sub>x=tutoring</sub> | 0.033  | 0.043 | 0.782   | 0.436   | 0.036       |               |
| Gender           | CE <sub>K=boys</sub>     | 0.015  | 0.076 | 0.196   | 0.846   | 0.016       | 0.070         |
|                  | CE <sub>K=girls</sub>    | -0.051 | 0.080 | -0.630  | 0.531   | -0.054      |               |
| Migration        | CE <sub>K=no</sub>       | -0.010 | 0.075 | -0.135  | 0.893   | -0.011      | 0.033         |
|                  | CE <sub>K=yes</sub>      | -0.042 | 0.073 | -0.567  | 0.572   | -0.044      |               |
| Pretest          | CE <sub>Z=(-1 SD)</sub>  | 0.031  | 0.060 | 0.525   | 0.601   | 0.033       | 0.093         |
|                  | CE <sub>Z=(+1 SD)</sub>  | -0.057 | 0.126 | -0.449  | 0.656   | -0.060      |               |
|                  | CE <sub>Z=(-2 SD)</sub>  | 0.079  | 0.107 | 0.742   | 0.462   | 0.084       | 0.191         |
|                  | CE <sub>Z=(+2 SD)</sub>  | -0.101 | 0.204 | -0.494  | 0.625   | -0.107      |               |

CE = conditional effect on reading literacy (PV) in Grade 7. The reported effects are adjusted for all covariates. We report unstandardized effect estimates (Est.), robust standard errors (SE), tests of estimates against zero (Est./SE and p-values), effects in standard deviation units of the outcome variables in the control groups (effect size) and differences of effect sizes (Δ effect size)

**Effects of extracurricular tutoring in German**

Concerning the effects of ETG (RQ3.1), first, we investigated average treatment effects (AE) as reported in Table 2. The results show no significant effect of ETG on students’ reading literacy, except for the unadjusted mean negative effect, which does not account for differences between covariates at the beginning of the assessment. However, when controlling for covariates—particularly for the reading literacy in the pretest and then for additional student characteristics—this negative effect turned smaller and was no longer significant. This shows that the pretest reading literacy was the most important covariate to control for baseline differences between the treatment groups.

Second, we considered subgroup-specific treatment effects (CE) in relation to the participation status, gender, migration background, and specific pretest values (RQ 3.2; see Table 3). Overall, the results show no significant effects of ETG. Nevertheless, slight patterns are discernible in relation to (non-significant) differences in the effect size estimates: (a) effects were higher for the subgroup of students currently participating in tutoring (i.e., conditional treatment effect on the treated compared to those who did not participate in tutoring [conditional treatment effect on the non-treated]), (b) boys benefited slightly more from tutoring than girls, (c) native German students benefited slightly more from tutoring than students with migration background, and (d) students with lower initial reading literacy in Grade 5 seem to have benefited more from ETG than students with higher reading literacy at the beginning. Although these trends are



**Table 4** Conditional treatment effects of tutoring in German for interactions with treatment status

| Grouping                | Effect                                | Est.   | SE    | Est./SE | p-value | Effect size | Δ effect size |
|-------------------------|---------------------------------------|--------|-------|---------|---------|-------------|---------------|
| Treatment and gender    | CE <sub>X=control, K=male</sub>       | 0.010  | 0.080 | 0.123   | 0.903   | 0.011       | 0.068         |
|                         | CE <sub>X=control, K=female</sub>     | -0.054 | 0.083 | -0.644  | 0.522   | -0.057      |               |
|                         | CE <sub>X=tutoring, K=male</sub>      | 0.059  | 0.055 | 1.077   | 0.285   | 0.063       | 0.073         |
|                         | CE <sub>X=tutoring, K=female</sub>    | -0.010 | 0.059 | -0.165  | 0.869   | -0.010      |               |
| Treatment and migration | CE <sub>X=control, K=native</sub>     | -0.015 | 0.078 | -0.192  | 0.848   | -0.016      | 0.066         |
|                         | CE <sub>X=control, K=migration</sub>  | -0.047 | 0.077 | -0.615  | 0.540   | -0.050      |               |
|                         | CE <sub>X=tutoring, K=native</sub>    | 0.055  | 0.048 | 1.142   | 0.256   | 0.059       | 0.067         |
|                         | CE <sub>X=tutoring, K=migration</sub> | -0.007 | 0.065 | -0.110  | 0.912   | -0.008      |               |
| Treatment and pretest   | CE <sub>X=control, Z=-1SD</sub>       | 0.030  | 0.063 | 0.478   | 0.634   | 0.032       | 0.097         |
|                         | CE <sub>X=control, Z=+1SD</sub>       | -0.061 | 0.127 | -0.482  | 0.632   | -0.065      |               |
|                         | CE <sub>X=tutoring, Z=-1SD</sub>      | 0.041  | 0.060 | 0.696   | 0.489   | 0.044       | 0.077         |
|                         | CE <sub>X=tutoring, Z=+1SD</sub>      | -0.031 | 0.134 | -0.234  | 0.816   | -0.033      |               |
|                         | CE <sub>X=control, Z=-2SD</sub>       | 0.077  | 0.112 | 0.687   | 0.497   | 0.082       | 0.189         |
|                         | CE <sub>X=control, Z=+2SD</sub>       | -0.101 | 0.204 | -0.494  | 0.625   | -0.107      |               |
|                         | CE <sub>X=tutoring, Z=-2SD</sub>      | 0.086  | 0.108 | 0.795   | 0.431   | 0.091       | 0.135         |
|                         | CE <sub>X=tutoring, Z=+2SD</sub>      | -0.041 | 0.153 | -0.269  | 0.789   | -0.044      |               |

Conditional effects (CE) on reading literacy (PV) in Grade 7 are reported for the non-treated and treated students (with tutoring in German in Grades 5, 6 or 7), while adjusting for all covariates. From the set of covariates (K), grouping variables (i.e., treatment variable [X]: tutoring vs. control, gender: male vs. female, migration: native vs. migration, pretest (Z): [- 2SD], [- 1SD], [+ 1SD] [+ 2SD]) are used and combined for obtaining the subgroup-specific effects. We report unstandardized effect estimates (Est.), robust standard errors (SE), tests of estimates against zero (Est./SE and p-values), effects in standard deviation units of the outcome variables in the control groups (effect size) and differences of effect sizes (Δ effect size)

identifiable from the comparisons of the effect size estimates and congruent to our expectations, it must be mentioned, that all effect estimates were not significantly different from zero and the effect sizes were very small (< |0.10|).

Finally, in a third step, we obtained even more detailed conditional treatment effects for the participating students (ETG) and non-participating students (no ETG) by combining the participation status with each grouping variable (see Table 4). The overall trend was the same as before, with the highest effect sizes for students currently participating in ETG, as well as that tutoring is slightly more beneficial for male students, for those that have a native German background, and those with lower initial reading literacy in Grade 5. Also, all effect estimates were not significantly different from zero and all effect sizes were smaller than |<0.10|; except for students who did not participate in ETG who had very high initial reading literacy. This subgroup of good readers had a small negative effect size of - 0.11 standard deviations of the outcome variable.

**Discussion**

Regarding ETG between Grades 5 and 7 (RQ1.1), our results revealed a low participation rate in the total sample (about 9%). This rate was lower than expected and lower compared to previous results (about 19%; Guill & Wendt, 2016; Guill et al., 2022), but closer to the participation rate for paid tutoring in German found for students in Grade 9 (about 3%; Hawrot, 2024). Similarly, the participation rate for ETR (about 4%) was much lower than expected based on earlier results (about 16%; Guill & Wendt, 2016). Accordingly, very few students overall participated in interventions, such as text comprehension exercises, that are recommended to promote reading literacy (Baye

et al., 2018; Edmonds et al., 2009; Souvignier & Antoniou, 2007). However, one reason for the variation in participation rates compared to previous results may be that parents do not know exactly whether or to what extent their children are currently participating in ETG or ETR as tutoring participation is often organized by schools in Germany. A second reason might be a difference in definitions of tutoring as mentioned above, for example, paid tutoring in Hawrot's study (2024) compared to all tutoring offers, whether paid or not, in this study.

Concerning participation rates in ETG and ETR of different subgroups of students (RQ1.2), our results revealed significant differences in the likelihood for participation: first, it was more likely to participate in ETG and also in ETR for students with lower cultural capital compared to contemporaries with higher cultural capital as previously shown by Guill et al. (2020b). Second, gender mattered in our study in contrast to findings by Hawrot (2024): according to Guill et al. (2022), the likelihood of participation in ETG and ETR was higher for boys compared to girls. Third, students with a migration background were more likely to participate in ETG and ETR compared to German students which is consistent with previous findings (Guill et al., 2020b; Hawrot, 2024) as well as suggestions of promoting students with migration background (Artelt et al., 2010; Diedrich et al., 2019; Weis et al., 2019). Fourth, our results showed a higher likelihood of participating in tutoring for students in non-academic schools than for students in academic schools which is consistent with previous studies (Hawrot, 2024). Fifth, in line with previous findings (Guill et al., 2020b; Hawrot, 2024), tutoring participants of ETG in our study tended to have weaker grades in Grade 5 compared to non-participating students. Nevertheless, only about one-quarter of these students participated in ETG. Finally, according to previous results (Tarelli et al., 2012; Weis et al., 2016) and recommendations (Artelt et al., 2010; Diedrich et al., 2019; Weis et al., 2019), many of the students for whom support is advised, particularly boys, appear to participate in ETG, fewer in ETR. About 20% of the students with poor reading literacy (scores one or two standard deviations below the mean)—in line with previous findings (Tarelli et al., 2012; Weis et al., 2016) representing about 16% of the total sample—participated in ETG, while good readers participated comparatively rarely. Nevertheless, it should be noted that the number of poor readers in need of support is presumably higher than the number of actual participants in tutoring, who, after all, make up only a very small percentage of the total sample (about 8.5%). The differences in our results compared to the participation rates reported of other studies could be due to the fact that we examined students between Grades 5 and 7, whereas other studies focused on students in other Grades. A further reason for inconsistent results could be differences of definitions of tutoring. However, participation rates in the more specific ETR are considerably lower than in ETG, thus the various subgroups are quite small and the effects of ETR cannot be estimated in our models as discussed in the following.

Concerning the characteristics of extracurricular tutoring in German (RQ2), our results indicated a high heterogeneity, particularly with respect to social settings, type of tutors, locations, and weekly scope: nevertheless, the weekly scope of tutoring in German (and possibly additional subjects) was similar for most students (75%), specifically, it was limited to one semester and only 1 or 2 h per week. Although it is not given exactly from the available data which subject each hour is spent on, the reported

amount suggests that the time spent for tutoring is within the range of times found to be effective in previous reading supporting programs (Holopainen et al., 2018; Souvignier & Antoniou, 2007). Compared with the weekly scope and duration of effective tutoring in various subjects (Mischo & Haag, 2002), the scope of ETG and ETR found was much lower. However, for future analyses, it would be helpful to assess exactly what time in tutoring is devoted to specific topics. Regarding the social settings of tutoring, the results showed that tutoring in German (and possibly additional subjects) for most students—mainly privately at home or school—was conducted individually rather than in small groups and rarely in large groups. Thus, the social settings of tutoring found were largely consistent with practical implications of previous studies, which found one-to-one (Baye et al., 2018; Suggate, 2016) and small group tutoring in reading (Almasi & Palmer, 2013; Holopainen et al., 2018) to be the most effective settings compared to large group tutoring. In addition, our results showed that tutoring in German (and possibly additional subjects) was most frequently provided by teachers, followed by (school) students and private persons. This result partly matches with practical implications of earlier studies suggesting tutoring in reading with trained tutors, peers or teachers to gain the highest effects on tutees' reading literacy compared to other types of tutors (Suggate, 2016). Since current data lack information on the extent to which tutoring (school) students and private persons have been trained for their tutor role, which has been shown to be an important prerequisite for progress in tutees' reading literacy, it is promising for future studies to pay more attention to this aspect. Finally, it should be noted critically that our data on characteristics of tutoring in German contain a large number of missings and, thus, provide no representative picture of extracurricular tutoring practice in Germany.

Finally, our analyses revealed no effect of students' participation in ETG on their reading literacy (RQ3.1). Nevertheless, using different sets of covariates for estimating the average effect of ETG showed that covariate adjustment is very important in this application, to control for baseline differences between the treatment groups. Without covariate adjustment, one would falsely obtain a large negative effect, indicating that students who participate in ETG substantially showed lower reading literacy afterward compared to students without participation. In contrast, when controlling for differences in the initial reading literacy of the students in Grade 5, this negative effect diminished, showing that on average, ETG had no significant effect on reading literacy in Grade 7. Additional covariates had little impact on this conclusion, but with more covariates, the average effect estimate was a bit closer to zero. Thus, our results were consistent with previous findings that showed no effects of participation in extracurricular activities or tutoring in German on students' achievement (Guill et al., 2022; Shulruf, 2010; Shulruf & Wang, 2013) and no effects of tutoring in German or mathematics on students' grades in respective subjects (Baker et al., 2001; Guill et al., 2020a, 2022). However, our results were inconsistent with previous findings that showed positive associations between students' achievement and their participation in extracurricular activities (e.g., Carbonaro & Maloney, 2019; Feldman & Matjasko, 2005; Lewis, 2004; Mischo & Haag, 2002). The fact that our results differ from previous findings may be due to various differences across studies, namely differences (a) in design (e.g., quasi-experimental groups and various defined interventions, Mischo & Haag, 2002), (b) in methodological procedures (e.g., cross-lagged panel models, Carbonaro

& Maloney, 2019), (c) in the conceptions of tutoring (e.g., tutoring as paid extracurricular support), and d) in different amount and intensity of tutoring. In particular with respect to the study by Mischo and Haag (2002), it is striking that the scope of the ETG (mostly about 1–2 h per week) found is much smaller than the scope of the reported tutoring for which effects were demonstrated (about 6 h weekly). Moreover, the results contradict our expectations of the effects of ETG, which were based on the assumption that the ETG would include interventions that have been shown to be effective in reading promotion programs (e.g., Almasi & Palmer, 2013; Baye et al., 2018; Edmonds et al., 2009; Holopainen et al., 2018; Souvignier & Antoniou, 2007; Suggate, 2016). The fact that the results do not support our assumption may also indicate that the relevant measures have not been implemented. Finally, another reason for our unexpected findings may be that the tutoring methods and exercises were examined from a too general perspective as discussed below.

In addition, we controlled for various covariates and examined differential effects for students of different subgroups (RQ3.2). However, these analyses showed no differential effects of ETG. In particular, we found no significant effects on reading literacy for specific subgroups of students who differed in terms of migration background, gender, cultural capital at home, German grades in Grade 5, school type, and initial reading literacy. Still, comparisons of the effect sizes obtained for the different subgroups suggest marginal differences in the effects of ETG for these students. Accordingly, higher effects on later reading literacy were present for the group of students that actually participated in ETG compared to the group of students that did not participate (effects on the treated and non-treated), for boys compared to girls, for students with a German background compared to students with a migration background, and for struggling readers compared to good readers. With respect to students with different levels of initial reading literacy, results showed the highest positive effects of ETG for struggling readers ( $-2SD$ ), second highest positive effects for poor readers ( $-1SD$ ), and slightly negative effects for good and very good readers. These effects are in the expected direction, but also very small and not significant. According to the expectations, ETG might help to compensate for achievement gaps and thus ultimately mitigate or avoid scissor effects found in students between Grades 5 and 7 (Pfost & Artelt, 2013). On the one hand, our expectations were based on previous results indicating strong effects of reading promotion programs (Baye et al., 2018; Edmonds et al., 2009; Souvignier & Antoniou, 2007), particularly for struggling readers. On the other hand, these expectations were based on earlier proposals to implement extracurricular tutoring for high-needs students, especially those with reading difficulties, boys, and students with a migration background (Artelt et al., 2010; Diedrich et al., 2019; Tarelli et al., 2012; Weis et al., 2016, 2019). However, the expected effects were not evident in our results. In addition to the overly general perspective on tutoring methods used in ETG mentioned above, another reason that expected effects were not revealed could be that the reading literacy measures used were too general. Both aspects are discussed in the following section.

### **Limitations of the study**

Several weaknesses might have limited the conclusions of our study. One limitation is that extracurricular tutoring in German and reading was investigated from a very general perspective, without distinguishing between individual support measures, e.g., exercises to improve reading speed or vocabulary, or others. On the one hand, we

chose this perspective in line with previous studies that suggest, for example, effects of subject-related activities on students' achievement (e.g., Hattie, 2009), or examined the effects of tutoring in German (e.g., Guill et al., 2020a, 2022; Mischo & Haag, 2002). On the other hand, from this general perspective, our data provided a sufficiently large sample to conduct analyses at this level. However, the required sample size was not available for more specific analyses of the effects of ETR. Finally, this critical feature of our study corresponds to the methodological issues reported above with respect to investigating the effects of extracurricular activities from a large-scale perspective, which was attributed to the fact that very heterogeneous methods were included (Bray, 2014; Shulruf, 2010; Shulruf & Wang, 2013). Accordingly, the investigation of ETG—as a summary of all methods for the promotion of competences in the subject German, including reading literacy—might have been one reason that we detected no substantial effects on reading literacy in our analyses. Particularly, the methods subsumed in ETG were still too diverse, and their effects, while potentially evident in improvements of specific skills—for example, knowing, understanding, and using words in oral or written texts—remained undetected under the radar of capturing reading literacy as an overarching measure. Therefore, for future analyses, it seems promising to precisely capture the respective methods implemented to improve specific student competences.

A second limitation of our study is the use of reading literacy as a target criterion. This rather distal measure cannot show possible effects of ETG on individual reading skills, such as vocabulary, reading speed, or other important basic skills of reading literacy (see also Fig. 1), and, thus, may underestimate the effects of ETG. Similar methodological problems are mentioned by Almasi and Palmer (2013) as well as by Souvignier and Antoniou (2007), who reported higher effects when study-specific measures—that is, designed by the researchers of the study with respect to the intervention goals and promoted competences—were used instead of standardized tests as a more proximal measure. Hence, according to their conclusions, it is useful to use both, researcher-designed and standardized tests, to measure the effects of support methods in terms of near and far transfer. Therefore, for future analyses it is promising to use additional instruments that capture the competences promoted in each method of tutoring, and thus, to focus on methods and appropriate target criteria, as illustrated at the same levels in Fig. 1 above.

Furthermore, it has to be mentioned critically that variables potentially influencing reading literacy, participation in extracurricular tutoring, or effects of ETG were not controlled for. On the one hand, various characteristics of students can be considered as potential influencing variables, for example, their reading behavior, reading self-concept, reading motivation (e.g., Heyne et al., 2023), as well as helplessness and subjective task values which predicted participation in reported studies (e.g., Hawrot, 2024). It is also conceivable that other qualitative features of extracurricular tutoring such as cognitive activation, social support, clear structure, or different topics (e.g., Baye et al., 2018; Guill et al., 2020a, 2022; Lewis, 2004) determine effects on students' performance. While some of these tutoring characteristics were not examined in our study, others were captured with a substantial nonresponse rate making it difficult to generalize our findings to German students. Therefore, future studies are recommended to systematically summarize the current state and specific characteristics of tutoring in Germany in representative samples.

## Conclusion

Finally, the study provided insights into participation rates, characteristics, and effects of extracurricular tutoring in German and reading in the everyday life of German students between Grades 5 and 7 from a longitudinal perspective. Against the background of existing cross-sectional large-scale studies that primarily examined reading literacy and derived corresponding suggestions for supporting many students on the one hand, our results for a 3-year perspective showed that students rarely use extracurricular tutoring in German and even less frequently in reading, which is particular true for struggling readers, for whom this offers are most appropriate. On the other hand, referring to studies on the characteristics of various extracurricular offers, our study also indicated that although the existing offers met some formal criteria based on practical implications from other studies, little is still known about the specific measures implemented. Finally, with reference to studies on the effects of extracurricular tutoring, such as reading promotion programs, our results revealed no effects of ETG on reading literacy, which could be due to various methodological features of the study, but also to the small amount of ETG observed.

However, based on these results and the current state of research, from our point of view it is premature to conclude that extracurricular tutoring has no effects on students' achievements. Instead of throwing the baby out with the bathwater in this way, it is crucial in future research to find out what (and to what extent) best promotes the reading literacy of whom. In order to implement this, three central desiderata for future research were derived from the results: there is a need for more detailed and representative surveys on extracurricular tutoring in German and reading with respect to (a) the learning-supporting characteristics, e.g., tutoring times; (b) the methods implemented, e.g., exercises to improve vocabulary, reading speed, text comprehension, or (question) strategies; and (c) the sub-skills of reading literacy in terms of additional target criteria of the methods implemented. For such more detailed data, our analysis approach of a differential effect analysis, in which we controlled for baseline differences as well as measurement error, can provide valuable insights into (group-specific) effects of any kinds of interventions. Our differential effect analyses imply that ETG can be more effective for certain students, thus, the analysis strategy and results raised new questions about for whom and how tutoring works best. With additional data, the differential effectiveness can be evaluated in more detail, or even at an individual level (e.g., Mayer et al., 2020).

Finally, for future practice, the present findings suggest that efforts should be stepped up to involve more students with support needs in extracurricular tutoring, especially those struggling readers and students with migration backgrounds who have rarely participated to date. In order to address their support needs, the implementation of evidence-based reading promotion programs is an option for which numerous evaluation studies have yielded promising results (Almasi & Palmer, 2013; Baye et al., 2018; Edmonds et al., 2009; Souvignier & Antoniou, 2007; Suggate, 2016). Implementing of such programs in extracurricular tutoring in reading suggests that participants can be effectively supported in their reading literacy development which finally is also expected to become evident in large-scale assessments.

## Appendix

See Tables 5 and 6.

**Table 5** Information on extracurricular tutoring in general, in German, and in reading

| Features                                      | Interview questions  |
|---|--|
| 1 Occurrence of private tutoring <sup>a</sup> | Does <name> currently receive private tuition? (private tutoring: all external educational, systematic, mainly regular support of students to overcome learning issues or to improve learning performance)   |
| 2 Subject of tutoring <sup>b</sup>            | And in what subjects is <name> receiving private tutoring? (asked for mathematics, <i>German</i> , English, French, Latin, physics, chemistry, biology, other)   |
| 3 Content of tutoring <sup>b</sup>            | What is the focus of your tutoring in German? (asked for spelling and writing, <i>reading and understanding texts</i> , writing texts, speaking and oral comprehension, grammar)   |
| 4 Scope of private tutoring <sup>c</sup>      | And how many hours in total per week does this private tutoring comprise in a normal school week? (in case of questions: a normal school week means not during the holidays, or at times when no private tutoring takes place for other reasons <sup>d</sup> ) |
| 5 Type of location <sup>d</sup>               | Where does <name> receive their private tutoring? (options are read aloud. If the tuition takes place in different venues: 'where does it mainly take place?')   |
| 6 Type of organization <sup>e</sup>           | And how is the tuition organized? (options are read aloud)   |
| 7 Type of mentor <sup>f</sup>                 | Who gives the tuition? (options are read aloud. If several persons are named: 'please tell me the name of the person who gives most of the private tutoring')  |

Questions in parents' interviews with different answering options: <sup>a</sup>"yes" (1), "no" (2), "irregular private tutoring" (– 20), "don't know" (– 98)

<sup>b</sup>"Not specified" (0), "specified" (1), "none of this" (– 20), "refused" (– 97), "don't know" (– 98)

<sup>c</sup>Number of hours (0–8, 10, 12, 15, 20, 23, 28, 30, 31, 34, 40, 45, 60, 90; for tutoring in several subjects hours are added), "irregular private tutoring" (– 20), "don't know" (– 98)

<sup>d</sup>"Privately, at your home" (1), "privately, but not at home" (2), "at a private tuition institute" (3), "at school" (4), "at a youth or community center" (5)

<sup>e</sup>"Individual tuition" (1), "in small groups, up to 5 students" (2), "in groups of more than 5 students" (3)

<sup>f</sup>"A qualified teacher" (1), "student" (2), "a school student" (3), "another private individual" (4); questions 1, 2, 3, 4 were used in waves 1, 2, 3, 4, 6; questions 5, 6, 7 were used in waves 2, 4 (see NEPS Network, 2021b)

**Table 6** Assessment of the students' migration background

| Codes for answers to the question "were you or one of your parents not born in Germany?" |                   |  |
|--|-------------------|--|
| Value  | Label             | Description  |
| 0  | Majority          | Target person and parents born in Germany, at most one grandparent (if any) born abroad  |
| 1  | 1st generation    | Target person born abroad and immigrated after the age of six  |
| 2  | 1.5th generation  | Target person born abroad and immigrated before the age of six   |
| 3  | 2nd generation    | Target person born in Germany and both parents born abroad   |
| 4  | 2.25th generation | Target person born in Germany and one parent born abroad; other parent born in Germany and both of that parent's parents born abroad |
| 5  | 2.5th generation  | Target person born in Germany, one parent born abroad; other parent born in Germany and one of that parent's parents born abroad     |
| 6  | 2.75th generation | Target person born in Germany, one parent born abroad; other parent born in Germany and neither of that parent's parents born abroad |
| 7  | 3rd generation    | Target person and parents born in Germany; all four grandparents born abroad   |
| 8  | 3.25th generation | Target person and parents born in Germany; three grandparents born abroad  |
| 9  | 3.5th generation  | Target person and parents born in Germany; two grandparents born abroad  |

Answering options in students' interviews (adapted from Olczyk et al., 2016). Students' answers were aggregated to *yes* (= 1; when options 1–6 were chosen) or *no* (= 2; when answers 7–9 were chosen)

### Abbreviations

|       |  |
|-------|--|
| + 1SD | Higher than one standard deviation above the mean                    |
| + 2SD | Higher than two standard deviations above the mean                   |
| − 1SD | Lower than one standard deviation below the mean                     |
| − 2SD | Lower than two standard deviations below the mean                    |
| AE    | Average treatment effect   |
| CE    | Conditional treatment effect   |
| ECA   | Extracurricular activities   |
| ETG   | Extracurricular tutoring in German                                   |
| ETR   | Extracurricular tutoring in reading                                  |
| IGLU  | Internationale Grundschul-Lese-Untersuchung                          |
| K     | Covariates   |
| KESS  | Study on competencies and attitudes of students                      |
| NEPS  | National Educational Panel Study                                     |
| PIRLS | Progress in International Reading Literacy Study                     |
| PISA  | Program for International Student Assessment                         |
| PV    | Plausible values   |
| RQ    | Research questions   |
| SC3   | Starting cohort 3 of the National Educational Panel Study in Germany |
| TIMSS | Trends in International Mathematics and Science Study                |
| X     | Treatment effects  |

### Acknowledgements

Not applicable.

### Author contributions

NH mainly contributed to the conception, theoretical framework, generation of questions, data aggregation, presentation and discussion of results and conclusions. MAS and TG mainly provided data analyses, including imputation and different effect analyses and prepared figures and tables.

### Funding

The NEPS data were collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). Since 2014, the NEPS survey has been conducted by the Leibniz Institute for Educational Pathways (LifBi) at the University of Bamberg in cooperation with a nationwide network. No funding was received for the present study, which was conducted using publicly available NEPS data.

### Availability of data and materials

The datasets analysed in the present study are from the National Educational Panel Study in Germany (NEPS; Starting Cohort 3; assessed from 2008 to 2013) and are available at <https://doi.org/10.5157/NEPS:SC3:10.0.0>. The data is freely accessible to all researchers, who must register for it. The computer code to reproduce all of the reported results has been made publicly available at the Open Science Framework (OSF). The study was not preregistered.

### Declarations

#### Ethics approval and consent to participate

The NEPS study is conducted under the supervision of the German Federal Commissioner for Data Protection and Freedom of Information (BfDI) and in coordination with the German Standing Conference of the Ministers of Education and Cultural Affairs (KMK) and—in the case of surveys at schools—the Educational Ministries of the respective Federal States. All data collection procedures, instruments and documents were checked by the data protection unit of the Leibniz Institute for Educational Trajectories (LifBi). The necessary steps are taken to protect participants' confidentiality according to national and international regulations of data security. Participation in the NEPS study is voluntary and based on the informed consent of participants. This consent to participate in the NEPS study can be revoked at any time.

#### Consent for publication

All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 23 March 2023 Accepted: 6 August 2024

Published online: 17 August 2024

### References

- Almasi, J. F., & Palmer, B. M. (2013). Reading comprehension programs. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 342–344). Routledge.
- Artelt, C., Naumann, J., & Schneider, W. (2010). Lesemotivation und Lernstrategien [Reading motivation and reading strategies]. 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). Waxmann.



- Baker, D. P., Akiba, M., LeTendre, G. K., & Wiseman, A. W. (2001). Worldwide shadow education: Outside-school learning, institutional quality of schooling, and cross-national mathematics achievement. *Educational Evaluation and Policy Analysis*, 23(1), 1–17. <https://doi.org/10.3102/01623737023001001>
- Baye, A., Inns, A., Lake, C., & Slavin, R. E. (2018). A synthesis of quantitative research on reading programs for secondary students. *Reading Research Quarterly*, 54(2), 133–166. <https://doi.org/10.1002/rrq.229>
- Blossfeld, H.-P., & Roßbach, H.-G. (Eds.). (2019). *Education as a lifelong process: The German national educational panel study (NEPS). Edition ZfE* (2nd ed.). Springer VS. <https://doi.org/10.1007/978-3-658-23162-0>
- Bray, M. (2014). The impact of shadow education on student academic achievement: Why the research is inconclusive and what can be done about it. *Asia Pacific Education Review*, 15, 381–389. <https://doi.org/10.1007/s12564-014-9326-9>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Routledge.
- Buhl, H. M., & Hilkenmeier, J. (2016). Bildung und Lesesozialisation im Elternhaus [Education and reading socialization at home]. In B. Kracke & P. Noack (Eds.), *Handbuch Entwicklungs- und Erziehungspsychologie* (pp. 1–17). Springer VS. [https://doi.org/10.1007/978-3-642-54061-5\\_10-1](https://doi.org/10.1007/978-3-642-54061-5_10-1)
- Carbonaro, W., & Maloney, E. (2019). Extracurricular activities and student outcomes in elementary and middle school: Causal effects or self-selection? *Socius Sociological Research for a Dynamic World*, 5, 1–17. <https://doi.org/10.1177/2378023119845496>
- Diedrich, J., Schiepe-Tiska, A., Ziernwald, L., Tupac-Yupanqui, A., Weis, M., McElvany, N., & Reiss, K. (2019). Lesebezogene Schülermerkmale in PISA 2018: Motivation, Leseverhalten, Selbstkonzept und Lesestrategiewissen [Reading-related characteristics of students in PISA 2018: Motivation, reading behaviour, self-concept and knowledge on reading strategies]. In K. Reiss, M. Weis, & E. Klieme (Eds.), *PISA 2018. Grundbildung im internationalen Vergleich* (pp. 81–109). Waxmann.
- Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research*, 79(1), 262–300. <https://doi.org/10.3102/0034654308325998>
- Feldman, A. F., & Matjasko, J. L. (2005). The role of school-based extracurricular activities in adolescent development: A comprehensive review and future directions. *Review of Educational Research*, 75(2), 159–210. <https://doi.org/10.3102/00346543075002159>
- Fischer, L., Rohm, T., Gnamb, T., & Carstensen, C. H. (2016). *Linking the data of the competence tests (NEPS survey paper no. 1)*. Leibniz Institute for Educational Trajectories, National Educational Panel Study. <https://doi.org/10.5157/NEPS:SP01:1.0>
- Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online*, 5(2), 50–79. <https://doi.org/10.25656/01:8424>
- Gomila, R., & Clark, C. S. (2022). Missing data in experiments: Challenges and solutions. *Psychological Methods*, 27(2), 143–155. <https://doi.org/10.1037/met0000361>
- Guill, K., Lüdtke, O., & Köller, O. (2020a). Assessing the instructional quality of private tutoring and its effects on student outcomes: Analyses from the German National Educational Panel Study. *British Journal of Educational Psychology*, 90, 282–300. <https://doi.org/10.1111/bjep.12281>
- Guill, K., Lüdtke, O., & Schwanenber, J. (2020b). A two-level study of predictors of private tutoring attendance at the beginning of secondary schooling in Germany: The role of individual learning support in the classroom. *British Educational Research Journal*, 46(2), 437–457. <https://doi.org/10.1002/berj.3586>
- Guill, K., Ömeroğulları, M., & Köller, O. (2022). Intensity and content of private tutoring lessons during German secondary schooling: Effects on students' grades and test achievement. *European Journal of Psychology of Education*, 37, 1093–1114. <https://doi.org/10.1007/s10212-021-00581-x>
- Guill, K., & Wendt, H. (2016). Außerschulischer Nachhilfeunterricht am Ende der Grundschulzeit [Extracurricular tutoring at the end of elementary school years]. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert, & D. Kasper (Eds.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 247–256). Waxmann.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hawrot, A. (2024). Do school-related factors affect private tutoring attendance? Predictors of private tutoring in Maths and German among German tenth-graders. *Research Papers in Education*, 39(1), 1–23. <https://doi.org/10.1080/02671522.2022.2089209>
- Heyne, N., Gnamb, T., Lockl, K., & Neuenhaus, N. (2023). Predictors of adolescents' change in reading literacy: The role of reading strategies, reading motivation, and declarative metacognition. *Current Psychology*, 43(26), 32061–32075. <https://doi.org/10.1007/s12144-022-04184-7>
- Holopainen, L. K., Kiuru, N. H., Mäkihönko, M. K., & Lerkkanen, M.-K. (2018). The role of part-time special education supporting students with reading and spelling difficulties from grade 1 to grade 2 in Finland. *European Journal of Special Needs Education*, 33(3), 316–333. <https://doi.org/10.1080/08856257.2017.1312798>
- Krannich, M., Jost, O., Rohm, T., Koller, I., Pohl, S., Haberkorn, K., Carstensen, C. H., Fischer, L., & Gnamb, T. (2017). *NEPS technical report for reading—Scaling results of starting cohort 3 for grade 7 (NEPS survey paper no. 14)*. Leibniz Institute for Educational Trajectories, National Educational Panel Study. <https://doi.org/10.5157/NEPS:SP14:2.0>
- Lechner, C. M., Bhaktha, N., Groskurth, K., & Bluemke, M. (2021). Why ability point estimates can be pointless: A primer on using skill measures from large-scale assessments in secondary analyses. *Measurement Instruments for the Social Sciences*, 3(2), 1–16. <https://doi.org/10.1186/s42409-020-00020-5>
- Lewis, C. P. (2004). *The relation between extracurricular activities with academic and social competencies in school age children: A metaanalysis*. Texas A&M University.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174. <https://doi.org/10.1007/BF02296272>
- Mayer, A., Dietzfelbinger, L., Rosseel, Y., & Steyer, R. (2016). The EffectLiteR approach for analyzing average and conditional effects. *Multivariate Behavioral Research*, 51(2–3), 374–391. <https://doi.org/10.1080/00273171.2016.1151334>

- Mayer, A., Zimmermann, J., Hoyer, J., Salzer, S., Wiltink, J., Leibing, E., & Leichsenring, F. (2020). Interindividual differences in treatment effects based on structural equation models with latent variables: An EffectLiteR tutorial. *Structural Equation Modeling*, 27, 798–816. <https://doi.org/10.1080/10705511.2019.1671196>
- McElvany, N., Becker, M., & Lüdtke, O. (2009). Die Bedeutung familiärer Merkmale für Lesekompetenz, Wortschatz, Lesemotivation und Leseverhalten [The importance of family characteristics for reading literacy, vocabulary, reading motivation, and reading behavior]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 41, 121–131. <https://doi.org/10.1026/0049-8637.41.3.121>
- Mischo, C., & Haag, L. (2002). Expansion and effectiveness of private tutoring. *European Journal of Psychology of Education*, 17(3), 263–273. <https://doi.org/10.1007/BF03173536>
- Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56(2), 177–196.
- NEPS Network. (2021a). *National educational panel study, scientific use file of starting cohort grade 5*. Leibniz Institute for Educational Trajectories (LifBi). <https://doi.org/10.5157/NEPS:SC3:10.0.0>
- NEPS Network. (2021b). *National Educational Panel Study, Codebook*. Leibniz Institute for Educational Trajectories (LifBi), Bamberg. <https://www.neps-data.de/Data-Center/Data-and-Documentation/Start-Cohort-Grade-5/Documentation>.
- OECD. (2009). *PISA 2009 assessment framework. Key competencies in reading, mathematics and science*. OECD. <https://doi.org/10.1787/9789264062658-en>
- OECD. (2016). *PISA 2015 Ergebnisse (Band I): Exzellenz und Chancengerechtigkeit in der Bildung*. PISA, W. Bertelsmann Verlag. <https://doi.org/10.1787/9789264267879-de>
- Olczyk, M., Will, G., & Kristen, C. (2016). *Immigrants in the NEPS: Identifying generation status and group of origin (NEPS survey paper no. 4)*. Leibniz Institute for Educational Trajectories. <https://doi.org/10.5157/NEPS:SP04:1.0>
- Pfost, M., & Artelt, C. (2013). Reading literacy development in secondary school and the effect of differential institutional learning environments. In M. Pfost, C. Artelt, & S. Weinert (Eds.), *The development of reading literacy from early childhood to adolescence. Empirical findings from the Bamberg BiKS longitudinal studies* (pp. 229–278). University of Bamberg Press.
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Robitzsch, A., Kiefer, T., & Wu, M. (2022). TAM: Test analysis modules. R package version 4.1-4. <https://CRAN.R-project.org/package=TAM>
- Rossee, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Scharl, A., Fischer, L., Gnams, T., & Rohm, T. (2017). *NEPS technical report for reading: Scaling results of starting cohort 3 for grade 9 (NEPS survey paper no. 20)*. Leibniz Institute for Educational Trajectories, National Educational Panel Study. <https://doi.org/10.5157/NEPS:SP20:1.0>
- Sengewald, M.-A., & Mayer, A. (2024). Causal effect analysis in non-randomized data with latent variables and categorical indicators: The implementation and benefits of EffectLiteR. *Psychological Methods*, 29(2), 287–304. <https://doi.org/10.1037/met0000489>
- Shulruf, B. (2010). Do extra-curricular activities in schools improve educational outcomes? A critical review and meta-analysis of the literature. *International Review of Education*, 56(5/6), 591–612. <https://doi.org/10.1007/s11159-010-9180-x>
- Shulruf, B., & Wang, G. Y. (2013). Extracurricular activities in secondary schools. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 324–326). Routledge.
- Sieben, S., & Lechner, C. M. (2019). Measuring cultural capital through the number of books in the household. *Measurement Instruments for the Social Sciences*, 1(1), 1–6. <https://doi.org/10.1186/s42409-018-0006-0>
- Souvignier, E., & Antoniou, F. (2007). Förderung des Leseverständnisses bei Schülerinnen und Schülern mit Lernschwierigkeiten – eine Metaanalyse. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 76, 46–63.
- Steinhauer, H. W., Altmann, C., Zinn, S., Goßmann, S., & Rässler, S. (2015). Sampling and weighting cohort samples in institutional contexts. *AstA Wirtschafts- und Sozialstatistisches Archiv*, 9, 131–157. <https://doi.org/10.1007/s11943-015-0162-0>
- Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, 49(1), 77–96. <https://doi.org/10.1177/0022219414528540>
- Tarelli, I., Valtin, R., Bos, W., Bremerich-Vos, A., & Schwippert, K. (2012). IGLU 2011: Wichtige Ergebnisse im Überblick [IGLU 2011: Important results at a glance]. In W. Bos, I. Tarelli, A. Bremerich-Vos, & K. Schwippert (Eds.), *Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 11–25). Waxmann.
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427–450. <https://doi.org/10.1007/BF02294627>
- Weis, M., Doroganova, A., Hahnel, C., Becker-Mrotzek, M., Lindauer, T., Artelt, C., & Reiss, K. (2019). Lesekompetenz in PISA 2018 – Ergebnisse in einer digitalen Welt [Reading literacy in PISA 2018—Results in a digital world]. In K. Reiss, M. Weis, & E. Klieme (Eds.), *PISA 2018. Grundbildung im internationalen Vergleich* (pp. 47–80). Waxmann.
- Weis, M., Zehner, F., Sälzer, C., Strohmaier, A., Artelt, C., & Pfost, M. (2016). Lesekompetenz in PISA 2015: Ergebnisse, Veränderungen und Perspektiven [Reading literacy in PISA 2015: Results, changes, and perspectives]. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme, & O. Köller (Eds.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (pp. 249–284). Waxmann.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.