Number of books at home as an indicator of socioeconomic status: Examining its extensions and their incremental validity for academic achievement

Birgit Heppt[1] · Melanie Olczyk[2] · Anna Volodina[3]

Abstract

The present study investigates the incremental validity of the traditional books-at-home measure and selected extensions (i.e., number of children's books and number of ebooks) for explaining students' academic achievement as measured by their academic language comprehension. Using multiple linear regressions, we additionally explore the role of the source of information (i.e., whether information is given by parents or children). Based on cross-sectional data of a German sample of 2353 elementary school children from Grades 2 through 4, we found that parents' information on the number of books and children's books contributed to students' academic language comprehension over and above parental occupation and education. Children's information on the number of books did not further increase the amount of explained variance, and the effects were smaller than those for parents' information. Yet, when investigated separately, both parents' and children's information on the number of books and children's books at home predicted students' academic language comprehension and mediated the relationship between more distal structural features of socioeconomic status (i.e., parents' occupational status and education) and the outcome variable. No effect emerged for the number of ebooks. Our findings point to the robustness of the traditional books-at-home measure when used in parent questionnaires.

Keywords Books-at-home measure · Validity · Socioeconomic status · Cultural capital · eBooks · Academic language

✉ Birgit Heppt
   birgit.heppt@hu-berlin.de

Extended author information available on the last page of the article

1 Introduction

Social inequalities in student achievement have been well documented across countries, domains, and age groups, with students from socioeconomically disadvantaged families typically performing below their classmates from more privileged families (Harwell et al., 2017; OECD, 2019; Sirin, 2005). Given this firm relation between socioeconomic background and academic achievement, reliable, valid, and–ideally–economic indicators of socioeconomic status (SES) are required for educational research. One such indicator that is frequently used in social research is the number of books at home (e.g., Eriksson et al., 2021; Jerrim & Micklewright, 2014; Mullis et al., 2012a, 2012b). This indicator has consistently been found to relate to students' academic achievement, as measured, for instance, by their reading comprehension (e.g., Eriksson et al., 2021; McElvany et al., 2009; Mullis et al., 2012b; Park, 2008), academic language proficiency (Heppt & Stanat, 2020; Volodina et al., 2020), or years of schooling (Evans et al., 2010). It also shows substantial relations to other indicators of SES (e.g., Eriksson et al., 2021), thus pointing to its general appropriateness for capturing students' socioeconomic background. Moreover, whereas other measures involve laborious coding (e.g., the coding of parents' current occupation for assessing the International Socio-Economic Index [ISEI]; Ganzeboom et al., 1992), the books-at-home measure does not require any coding, making it a very economical and easy-to-use indicator that is particularly helpful for use in large survey designs or large-scale assessments. However, despite its widespread use, only a few studies have aimed to investigate and possibly improve this measure's quality (e.g., Engzell, 2018; Sieben & Lechner, 2019). Thus, it is unclear whether the traditional books-at-home measure holds as a valid indicator of SES over other commonly used measures or whether its incremental validity can be increased by including important extensions, such as the number of children's books and the number of ebooks. As the books-at-home measure is frequently administered to children, we further explore potential differences in its predictive power depending on whether parents or children share the information. This question is of high methodological and research-practical relevance because participation rates are typically higher for students than for parents in school-based assessments (cf. Engzell & Jonsson, 2015). The present study aimed to investigate these questions using academic language proficiency as an indicator of academic achievement.

2 SES and student achievement

Social inequalities in student achievement have been reported for a range of performance indicators in different domains, such as reading, mathematics, and science (e.g., Eriksson et al., 2021; OECD, 2019; Sirin, 2005). Theoretical explanations of these inequalities often draw on resource- and investment-oriented approaches (e.g., Bourdieu, 1986; Conger & Donnellan, 2007; Erikson &

Jonsson, 1996). These approaches basically rely on the assumption that families from different socioeconomic backgrounds differ in the amounts and kinds of resources they can invest in their children and that these varied investments, in turn, result in social inequalities in student achievement. The literature typically distinguishes between three types of resources: (1) economic resources, such as household income or wealth (Hällsten & Thaning, 2021), (2) social resources, including social networks, that provide access to support and valuable information, and convey (educational) norms and values (Carbonaro, 1998; Coleman, 1988) and (3) cultural resources or cultural capital. According to Bourdieu (1986), cultural capital can be subdivided into incorporated, institutionalized, and objectified cultural capital. Whereas incorporated cultural capital refers to a person's long-lasting dispositions, such as attitudes and preferences, institutionalized cultural capital denotes the certificates and academic titles that are obtained through formal education. Finally, objectified cultural capital includes physical cultural goods, such as books or works of art (Bourdieu, 1986; Sieben & Lechner, 2019). While some researchers highlight the role of cultural resources as opposed to economic resources, thus differentiating conceptually between SES and cultural capital (Evans et al., 2010; Park, 2008), others propose broader conceptualizations of SES that comprise cultural resources, such as the number of books at home (see Engzell, 2018, for an overview; Eriksson et al., 2021; Hanushek & Woessmann, 2011; Harwell et al., 2017). In line with this latter conceptualization, we use SES as an umbrella term, encompassing various economic, educational, and cultural resources that are often used for explaining social inequalities in student achievement.

Regarding their relations to students' academic achievement, it must be considered that the different resources are closely intertwined and can–at least to some extent–be converted into one another. Institutionalized cultural capital, for example, as reflected in a person's educational qualifications, can be transferred into economic resources. These enable parents to support the academic achievement of their children by providing them with better learning conditions (e.g., workplace) and learning materials in their homes (see Conger & Donnellan, 2007, for an overview). However, despite this overlap, the different resources refer to different aspects of a family's SES. They are therefore likely to be differentially related to academic achievement and, thus, cannot be used interchangeably (e.g., Bukodi & Goldthorpe, 2013; Eriksson et al., 2021; Hällsten & Thaning, 2021). In line with these considerations, studies incorporating different family resources typically report the respective indicators of SES to be substantially correlated while still highlighting that each indicator captures unique aspects of a family's socioeconomic background (Bukodi & Goldthorpe, 2013; Erola et al., 2016). Current research further highlights that various aspects of SES differ in the degree to which they are transferrable from one generation to the next. Unlike education and occupation, which are closely tied to individuals, wealth is only loosely related to an individual's effort and labor market success and, thus, can be passed on relatively easily and directly from parents to their children (Hällsten & Thaning, 2021). Yet, wealth's independent effect on student achievement seems negligible, at least in the major economies, whereas

particularly pronounced effects have been reported for parents' occupational status and the number of books at home (Eriksson et al., 2021).

In regard to the number of books at home, resource-oriented approaches, such as the family investment model (Conger & Donnellan, 2007), classify them as part of a family's cultural resources and stress their importance as potential learning stimulation at home (cf. Mullis et al., 2012a). As such, they may foster students' reading motivation and their interest in written culture (cf. McElvany et al., 2009) and form an important basis for stimulating learning interactions among parents and children. Specifically, they may serve as drivers for engaging in joint book reading activities, storytelling, talking about reading experiences, and further home literacy activities (Burgess et al., 2002; Eriksson et al., 2021; Grolig et al., 2019; Gustafsson et al., 2011). The number of books at home can thus be conceived of as a proximal process-oriented feature of SES that should be more closely linked to student learning outcomes than more distal structural features of SES, such as parental occupation status, parental education, or family income. Confirming this relational pattern, previous research found that the number of books (McElvany et al., 2009) and, more broadly, learning materials and learning stimulation available at home (Baydar & Akcinar, 2015) mediated the association between other features of SES (i.e., economic well-being and education) and children's language development. McMullin et al. (2020) found that the mediating role of the number of children's books between different operationalizations of a family's SES (e.g., mothers' education, family income) and children's language development (i.e., expressive vocabulary) was even more pronounced than that of home learning activities, such as reading to the child or helping the child learn songs, poems, or nursery rhymes (for similar findings, see Martin & Mullis, 2013). The number of books at home therefore seems to capture unique aspects of a family's SES that are key in explaining students' academic achievement in general and their language development in particular. Moreover, it has been found to mediate the relationship between more distal structural SES-features, such as parents' occupational status and education as well as student achievement.

That said, it is important to bear in mind that prior research points to reciprocal relations between literacy activities and language-related learning outcomes, such as reading comprehension (e.g., Harlaar et al., 2007). In a meta-analysis of 99 studies, for instance, Mol and Bus (2011) identified "an upward spiral of causality" (p. 267) of print exposure and language proficiency (i.e., oral language skills such as vocabulary, basic reading and spelling skills, and reading comprehension). Thus, children with a more stimulating reading environment in their homes tend to experience larger growth in oral language and reading comprehension. In turn, the better their reading comprehension, the more they engage in literacy-related activities (Mol & Bus, 2011). Such reciprocal effects and, consequently, reverse causality (cf. Leszczensky & Wolbring, 2019), may also come into play when investigating the relationships between the number of books at home and students' academic achievement.

3 SES and students' academic language proficiency

In recent years, a growing number of studies have pointed to social inequalities in students' academic language proficiency (Uccelli et al., 2019; Volodina et al., 2020, 2021a). Academic language is typically conceived of as the language register of schooling that students must master to participate in classroom discourse, understand textbooks, and accomplish assignments (e.g., Bailey, 2007; Snow, 2010). No unequivocal boundary exists between everyday language, which is used in daily routines and interactions, and the school-based register of academic language; different disciplines also involve different variations of the academic register (e.g., use of subject-specific vocabulary; Snow, 2010). Nevertheless, researchers generally agree that there are certain common language characteristics that are much more prevalent in school-based discourse than in everyday settings and that result in concise, factual, and information-dense texts in both oral and written forms (e.g., Bailey, 2007; Snow, 2010). Among these common characteristics are lexical features such as an often abstract and ambiguous academic vocabulary (e.g., *to determine, assumption*), an increased use of nominalizations, or the occurrence of complex connectives with very specific meanings (e.g., *although, subsequently*). These connectives, in turn, contribute to the construction of syntactically complex sentences (e.g., containing embedded subordinate clauses), which form typical grammatical features of academic language (cf. Schleppegrell, 2004; Volodina et al., 2021b).

School textbooks are replete with various features of the academic register. For instance, information in school textbooks is often presented in an abstract way (Achugar & Schleppegrell, 2005; Berendes et al., 2018). As early as elementary school, textbooks in mathematics, science, and social studies expose children to relatively high amounts of academic vocabulary (Fitzgerald et al., 2020, 2022). Moreover, coherence in textbooks is often achieved through the use of connectives (e.g., *perhaps, consequently, even;* Rodgers, 1974) while extensive explanations are circumvented through the use of short, albeit complex, noun phrases (Berendes et al., 2018).

Over the past few years, several studies have shown that students' general academic language proficiency is more closely related to academic achievement than more basic language skills and that academic language predicts student performance, even when controlling for general vocabulary or sentence comprehension (e.g., Schuth et al., 2017; Volodina et al., 2021b). These relational patterns have been shown across countries and for different age groups and domains (e.g., Meneses et al., 2018; Volodina et al., 2021b). The findings thus confirm academic language proficiency as a crucial precondition for school success and therefore emphasize its importance as an outcome variable when examining social inequalities in student achievement.

In investigating the relations between students' family background and their academic language proficiency, previous research drew on a variety of SES indicators, including the number of books at home (e.g., Heppt & Stanat, 2020; Volodina et al., 2020). While most of these studies identified the number of books at

home as a significant predictor of students' academic language comprehension, even when considering other measures such as parents' education and/or parents' occupational status, they typically relied only on the number of printed books. That is, the role of important extensions of the traditional books-at-home measure, which might contribute to its incremental validity, has not been investigated systematically.

4 Possible extensions of the traditional books-at-home measure

Studies that draw on the books-at-home measure as an indicator of SES usually use a single item that asks for the number of books at home (e.g., Evans et al., 2010; PISA; OECD, 2014). As this item is frequently accompanied by verbal descriptions and/or illustrations specifying how many books would fit on a bookshelf, it can be assumed that they primarily refer to printed books. While newspapers, magazines, and textbooks are sometimes explicitly excluded from the book count, this is usually not the case for the number of children's books. Given this blurring in item formulation, assessments should certainly benefit from clearer specifications of which sorts of books should be included in the estimate. Moreover, specific kinds of books that might be particularly important for the validity of the books-at-home measure should be assessed separately. Two such indicators are the number of children's books and the number of ebooks.

4.1 Number of children's books

The rationale for additionally assessing the number of children's books is twofold. First, considering the role of the books at home as potential learning stimulation, it can be assumed that for school-aged children, learning interactions are more likely to evolve around children's books than around parents' books. Second, it might also be easier for children to estimate their own books than their parents' books; therefore, this indicator may be more convenient for inclusion in student questionnaires than the traditional books-at-home measure (cf. Pagel, 2016). While previous research reported low agreement between parents' and children's information on the number of (parents') books (e.g., Engzell, 2018; Jerrim & Micklewright, 2014), information on the agreement is not available for the number of children's books since this measure is typically not assessed in student questionnaires.

A few studies have assessed both the number of parents' books and the number of children's books, but the independent effects of both indicators on academic achievement have typically not been singled out. That is, information on the number of parents' books and the number of children's books have usually been collapsed into a single variable, which was then used as a predictor of student achievement. While these studies confirm the mediating role of the number of books between parental education and performance in reading, science, and mathematics (Gustafsson et al., 2011; McElvany et al., 2009), they do not allow for an evaluation of the incremental validity of the number of children's books over the number of parents' books. As

an exception, an unpublished master's thesis conducted in Germany revealed that parents' information on the number of children's books contributed significantly to the explanation of students' reading comprehension, arithmetic skills, and academic language comprehension over and above the number of parents' books at home (Pagel, 2016). This finding supports the aforementioned assumption that the number and use of children's books, even more so than the number of parents' books, are highly relevant as learning stimulation during childhood. We therefore assume that the number of children's books predicts students' academic language comprehension over and above the number of parents' books.

4.2 Number of ebooks

Considering the ongoing process of digitalization that is accompanied by an enhanced access to ebooks and other digital devices, the traditional books-at-home measure may no longer be an accurate indicator of a family's cultural resources (Schwippert, 2019). It therefore seems reasonable to additionally ask for the number of ebooks to increase the incremental validity of the books-at-home measure (Pagel, 2016).

Although this argument seems traceable at first sight, empirical findings indicate that, at least in Germany, printed books are still used much more frequently than ebooks. While the percentage of people who sometimes read ebooks increased from 21% in 2013 to 30% in 2020, the share of people who sometimes read printed books grew from 74 to 81% during the same time span (Statista, 2021). This gap is even more pronounced for children: whereas approximately 70% of all children aged 4–13 use printed books at least once a week, the share of children who regularly use ebooks is below 10% at all ages (Kinder-Medien-Studie, 2018; for corresponding findings from the US, see Rideout, 2014). In addition to this rather low spread of ebooks, research suggests that parents and children tend to use ebooks and printed books differently. Specifically, they spend less time on content-related talk when engaging in joint ebook reading compared to joint reading of printed books, resulting in limited text comprehension in children (Krcmar & Cingel, 2014; Ross et al., 2016).

Taken together, both the relatively low availability of ebooks and the possibly less effective interaction patterns associated with joint ebook reading may limit the predictive validity of the number of ebooks for students' academic performance. In line with these considerations, using data from 2014, the previously mentioned master's thesis found no widespread use of ebooks and no incremental validity of the number of ebooks in the prediction of academic achievement (Pagel, 2016). However, to the best of our knowledge, virtually no published studies have investigated the predictive validity of the number of ebooks as an indicator of SES. The present study addresses this lacuna by examining whether the number of ebooks contributes to the prediction of students' academic language comprehension over and above the number of printed books. Given the relatively low proliferation and use of ebooks (Rideout, 2014; Statista, 2021), we expect the effects to be smaller than those for children's books.

5 Research questions and hypotheses

The current study aims to investigate the incremental validity of the books-at-home measure and selected extensions in predicting students' academic achievement. Considering that the number of books at home are frequently included in student questionnaires in national and international studies, we additionally explore the role of the information source, that is, whether parents or students share information. Specifically, we address the following research questions:

(1) Does the number of books at home contribute to students' academic language comprehension, when controlling for other common indicators of family SES (incremental validity)? Based on theoretical accounts and previous findings that highlighted the distinctness of different SES measures (Bukodi & Goldthorpe, 2013) and the proximity of the number of books at home to other process-oriented aspects of the home literacy environment (Gustafsson et al., 2011; McElvany et al., 2009), we expect the number of books at home to be significantly related to students' academic language comprehension even when considering important control variables (i.e., students' grade level and language background) and other indicators of family SES (i.e., parental occupation status and parental education).

(2) Is the incremental validity of the number of books at home increased by extensions of the traditional books-at-home measure (i.e., number of children's books and number of ebooks)? Assuming that children's books may be a more relevant resource for student learning and stimulating interactions than parents' books (cf. Pagel, 2016), we hypothesize that the number of children's books predicts students' academic language comprehension over and above the number of parents' books. Comparably smaller effects are expected for the number of ebooks, given their relatively low proliferation and use (Rideout, 2014; Statista, 2021).

(3) Are the effects of parental occupation status and parental education on students' academic language comprehension mediated by the number of books and children's books at home? Whereas parental occupation status and education have previously been described as distal structural features of SES, the number of books at home can be conceived of as a proximal process-oriented feature of SES that should be more closely related to students' learning outcomes in general and to language development in particular (McElvany et al., 2009). Replicating prior research that identified the number of books as mediator of the relation between more distal structural features of SES and students' academic achievement (Gustafsson et al., 2011; McElvany et al., 2009; McMullin et al., 2020; Myrberg & Rosén, 2009), we expect both the number of books at home and the number of children's books to mediate the relationship between distal structural features of SES (i.e., parental occupation status and parental education) and students' academic language comprehension.

(4) Does the predictive value of the number of books at home differ by information source (parents vs. children)? As previous research pointed to elementary school children's limited ability to estimate amounts (cf. Harel et al., 2007)

and found rather low consistency on parents' and children's information on the number of (parents') books (e.g., Engzell, 2018; Jerrim & Micklewright, 2014), we assume parents' estimates on the number of books to more accurately predict students' academic language comprehension than children's information. Yet, effects might be attenuated for the number of children's books, the number of which might be more easily accessible for children than for parents.

6 Materials and methods

6.1 Study information and sample

Data were drawn from the project "BiSpra-Transfer: Development and validation of a standardized test instrument" (German: *BiSpra-Aufgaben: Weiterentwicklung zu einem diagnostisch nutzbaren Testinstrument und Prüfung der Sensitivität für Fördereffekte*), which aimed to standardize three measures for assessing the academic language comprehension of primary school children (Heppt et al., 2020; Weinert et al., 2020). The study implemented a cross-sectional design and sampled students attending Grade 2, 3, or 4. Data collection took place in six German federal states in November and December 2017. We randomly sampled full classrooms for participation in the study and the overall sample consisted of 3778 primary school children from three student cohorts (for a more detailed description of the sampling procedure, see Heppt et al., 2020). By design, each student completed two of the three academic language measures. In each school, a coordinating teacher completed a student tracking list and provided information including students' grade level and gender. We additionally used a student questionnaire and a parent questionnaire to collect information on students' family background (i.e., language background, parental education and occupation) and the number of books at home. Schools, students, and parents participated in the study voluntarily and we obtained parents' written consent for their children's participation.

The analyses presented below are based on a subsample of 2353 students from Grades 2 ($n = 828$ students from 58 classrooms), 3 ($n = 780$ students from 56 classrooms), and 4 ($n = 745$ students from 50 classrooms) who were administered the measure on listening comprehension of academic language at the text level (*BiSpra-Text*) and for whom information on their language background (German monolingual, bilingual, German as a second language) was available. Students were 8.45 years old on average ($SD = 1.03$), and half of them were male. As the BiSpra-Transfer project aimed to provide testing norms for both German monolingual students and dual language learners (DLLs; for more detailed descriptions, see Heppt et al., 2020), 31% of the students in the sample simultaneously acquired German and another language (bilinguals), and 23% of the students started learning German when they were 3 years old or older (students with German as a second language).

6.2 Measures

6.2.1 Parents' information on the number of books at home

Parents indicated (1) the number of printed books at home (excluding ebooks, magazines, newspapers, and children's books), (2) the number of children's books (excluding magazines or newspapers), and (3) the number of ebooks. All three items were answered on a 5-point scale. For the number of printed books and the number of children's books, the five categories were "0–10" (1), "11–25" (2), "26–100" (3), "101–200" (4), and "more than 200" (5). To facilitate estimation, each of these categories was visualized by a bookshelf that included the respective number of books (cf. Richter et al., 2014). The item on the number of ebooks included the categories "0–10" (1), "11–25" (2), "26–50" (3), "51–100" (4), and "more than 100" (5). No visual aid was provided for this item.

6.2.2 Children's information on the number of books at home

Children indicated (1) the number of printed books at home (excluding magazines, newspapers, and their own books) and (2) the number of children's books (excluding magazines, newspapers, and schoolbooks). The items were answered on the same 5-point scale as the items in the parent questionnaire, and the five categories were illustrated by the same pictures of bookshelves. The descriptions of the five categories were slightly more detailed than in the parent questionnaire to make the connection between the pictures and categories more pellucid. Thus, the five categories were as follows: "none or only very few (0–10 books)" (1), "enough to fill one board in a shelf (11–25 books)" (2), "enough to fill a shelf (26–100 books)" (3), "enough to fill two shelves (101–200 books)" (4), and "enough to fill three or more shelves (more than 200 books)" (5). Children could read the descriptions in their questionnaires, and test administrators also read them out loud to avoid any comprehension difficulties.

6.2.3 HISEI

Furthermore, we asked the parents to report their current occupation. This information was coded according to the International Standard Classification of Occupations (ISCO-08; International Labour Office, 2012) and subsequently transformed into the ISEI (Ganzeboom et al., 1992). The ISEI is a standardized SES measure incorporating the income and necessary education level of a given occupation. It ranges from 10 to 90, with low values indicating occupations associated with low SES (e.g., unskilled workers) and high values indicating occupations associated with high SES (e.g., professors). The HISEI, that is, the highest ISEI among both parents was used for the current analyses. When information was available for only one parent, this information served as the HISEI.

6.2.4 Parental education

Parents indicated their highest level of schooling and vocational qualifications; we used this information to assess the families' educational background. Parents' information was first coded in line with the International Standard Classification of Education (ISCED-97; OECD, 1999), ranging from 0 (preprimary level of education) to 6 (second stage of tertiary education), and then transformed into the number of years of education (OECD, 2009). The so-called PARED varies between 0 and 18 years. We entered the highest PARED among both parents into our analyses. Again, when parents provided information for only one person, we used this information.

6.2.5 Academic language comprehension

The test instrument for assessing elementary school children's academic language comprehension in German (*BiSpra 2–4)* comprises three subtests, focusing on (1) listening comprehension of academic language at the text level (*BiSpra-Text*), (2) comprehension of connectives (*BiSpra-Sentence*), and (3) comprehension of cross-subject academic vocabulary (*BiSpra-Word;* Heppt et al., 2020). The following sections refer to *BiSpra-Text,* which broadly covers various academic language demands and can therefore be conceived of as a measure of general academic language comprehension, while the other two subtests assess specific language components.[1]

   BiSpra-Text has previously been described in detail (e.g., Heppt & Stanat, 2020; Heppt et al., 2020); therefore, we provide only a short description here. The measure assesses primary school students' ability to understand information and draw inferences from short listening comprehension texts that include a variety of lexical (e.g., cross-subject academic vocabulary, connectives) and grammatical (e.g., constructions in passive voice, syntactically complex sentences) features of the academic language register (cf. Bailey, 2007). The texts are entirely fictional and include made-up words that can be understood from the context, thus limiting undesirable effects of differences in prior content knowledge on students' test performance. *BiSpra-Text* includes three slightly different test versions for Grades 2, 3, and 4, each consisting of eight listening comprehension texts with four to seven linguistically simple yes/no questions. While some texts and items occur in only one test version, others appear across two or three test versions. Texts and items are presented auditorily via CD (for detailed test descriptions and sample items, see Heppt et al., 2020; Heppt & Stanat, 2020). Each of the three test versions' reliabilities was good in the present sample (Grade 2: $n_{\text{items}} = 38$, $\alpha = 0.83$; Grade 3: $n_{\text{items}} = 42$, $\alpha = 0.82$; Grade 4: $n_{\text{items}} = 44$, $\alpha = 0.83$).

---

[1] As a robustness check, we additionally conducted the same set of analyses for *BiSpra-Sentence* and *BiSpra-Word*. Analyses for these two measures showed a very similar pattern of results as those for *BiSpra-Text* and are reported in Tables S1 to S4 in the Supplemental Material.

6.3 Analytic procedure

Given the slightly different item sets for Grades 2, 3, and 4 for students' academic language comprehension, sum scores would not result in meaningful estimates of students' achievement across grades. We therefore jointly scaled students' item responses across all three grades using the R package TAM (Robitzsch et al., 2021). In the following analyses, we used weighted maximum likelihood estimates (Warm, 1989) as ability scores for *BiSpra-Text*.

The amount of missing data varied considerably among the study variables, ranging from 0% for students' academic language comprehension to 38% for the number of ebooks (Fig. S1 in the Supplemental Material). These differences can be partly explained by the different response rates for test booklets (100%), student questionnaires (98%), and tracking lists (100%) on the one hand and parent questionnaires (77%) on the other hand. To handle these missing data, we applied multiple imputation using the R package MICE (van Buuren & Groothuis-Oudshoorn, 2011), which generated 35 complete datasets. In addition to the study variables, the imputation model included a number of auxiliary variables that were substantially correlated with the study variables (e.g., students' grades in different subjects, perceived need for language support as indicated by the teachers) with the purpose of increasing the likelihood of a *missing at random mechanism* (cf. Collins et al., 2001).

To answer our research questions, we subsequently specified a number of multiple linear regression models in M*plus* Version 8.4 (Muthén & Muthén, 1998–2019). The mediating role of the number of books was investigated by including them as additional dependent variables and by using the "model indirect" command. We accounted for the data's nested structure (students in classes) with the option "type = complex", thus yielding robust standard errors. To integrate the analyses' results of the 35 datasets, we used the option "type = imputation" (Muthén & Muthén, 1998–2019).[2] Given the interrelatedness among the predictor variables, particularly among the SES indicators (see Sect. 7.1), we additionally checked for multicollinearity by determining the variance inflation factors (VIF). Pooled VIFs across all 35 datasets were obtained by specifying the respective regression models in STATA 17 and using *mivif* (Klein, 2011). While VIFs above 5 are typically considered as indicating substantial multicollinearity (Chatterjee & Simonoff, 2013; Hair et al., 2010; James et al., 2013), some authors propose lower thresholds of VIF ≥ 2.5 (Johnston et al., 2018). The VIFs varied from 1.10 to 2.61 across models and predictors.[3] Multicollinearity leads to increased standard errors and is therefore particularly problematic in smaller sample sizes (e.g., Hair et al., 2010). Compromising effects due to multicollinearity are thus unlikely to occur in the present analyses.

---

[2] When using the option "type = imputation" in M*plus*, parameter estimates are automatically averaged over the set of analyses. Pooled standard errors are based "[…] on the average of the standard errors over the set of analyses and the between analysis parameter estimate variation" (Muthén & Muthén, 1998–2019, p. 8).

[3] The VIFs range from 1.09 to 2.46 in the models predicting *BiSpra-Sentence* and from 1.09 to 2.41 in the models predicting *BiSpra-Word*.

Table 1 Descriptive Statistics and Manifest Correlations of the Study Variables

Number of measure	Measure	%	M	SD	1	2	3	4	5	6	7	8	9	10
1	Gender[a] (female)	50.10			–									
2	Grade				.05*	–								
	Grade 2	35.20												
	Grade 3	33.10												
	Grade 4	31.70												
3	Language background[b]				.00	.01	–							
	German monolingual	46.20												
	Bilingual	30.80												
	GSL	22.90												
4	HISEI[c]		44.61	21.01	-.03	-.05*	-.22*	–						
5	Parental education[d]		12.95	4.22	-.01	.00	-.22*	.59*	–					
6	Number of books (parents)[e]		2.92	1.30	-.01	-.02	-.31*	.53*	.47*	–				
7	Number of children's books (parents)[e]		2.61	1.05	-.01	-.01	-.32*	.47*	.44*	.73*	–			
8	Number of ebooks (parents)[f]		1.42	0.95	-.01	-.01	-.07*	.23*	.19*	.27*	.25*	–		
9	Number of books (children)[e]		2.89	1.40	-.03	-.07*	-.19*	.35*	.30*	.51*	.47*	.13*	–	
10	Number of children's books (children)[d]		2.50	1.19	-.01	-.06*	-.14*	.25*	.22*	.38*	.41*	.12*	.61*	–
11	Academic language comprehension		0.01	0.96	-.02	.25*	-.32*	.39*	.34*	.41*	.40*	.14*	.26*	.22*

GSL: German as a second language

[a] 0 = male, 1 = female

[b] 0 = German monolingual, 1 = bilingual student, 2 = student with German as a second language

[c] Highest International Socio-Economic Index of Occupational Status

[d] Parental education level (PARED) as indicated by the highest number of education years of both parents, which is based on the highest ISCED-level of both parents

[e] Assessed with a Likert scale, ranging from 1 (0–10 books) to 5 (more than 200 books)

[f] Assessed with a Likert scale, ranging from 1 (0–10 ebooks) to 5 (more than 100 ebooks). *p < .05

The analyses' code files in M*plus* and STATA can be found on OSF: https://osf.io/4vumy/?view_only=912d9aafc03d4d3080b4aff359195c54

7 Results

7.1 Descriptive results

Descriptive statistics and correlations of the study variables are displayed in Table 1. While the different SES indicators were all substantially and positively correlated, we found particularly high correlations for the number of books and children's books when the same person (i.e., parents or children) assessed them. Relatively small correlations emerged between the number of ebooks and all other indicators of family SES. Moreover, all indicators were positively correlated with academic language comprehension. In line with previous findings (e.g., Heppt & Stanat, 2020; Heppt et al., 2020; Volodina et al., 2020), we found a positive association between students' grade level and academic language comprehension. Students' language background was negatively related to academic language comprehension (e.g., Heppt et al., 2020; Volodina et al., 2020), indicating that DLLs performed below their monolingual peers, and to all indicators of SES. We therefore controlled for grade level and language background in the regression analyses. Gender was uncorrelated to academic language comprehension.

Using nonimputed data, we determined parent–child agreement for the number of parents' books and children's books and examined the distributions of the different books-at-home measures by source. Parent–child agreement was rather low (for an evaluation of kappa coefficients, see Cicchetti, 1994) within grades and overall for the number of both parents' books (Grades 2 to 4: $0.13 \leq \kappa \leq 0.24$; overall: $\kappa = 0.20$) and children's books (Grades 2 to 4: $0.11 \leq \kappa \leq 0.20$; overall: $\kappa = 0.14$). Distributions showed a substantially higher share of missing values for books-at-home measures included in the parent questionnaire than for those included in the student questionnaire (Fig. S1 in the Supplemental Material). Remarkably, the number of missing values was particularly large for the number of ebooks (38%), and almost 47% of the parents indicated that they had 0–10 ebooks compared to 11% who chose this category for the number of printed books and children's books.

7.2 Multiple linear regressions for explaining academic language comprehension

We performed a series of multiple linear regressions to explore the importance of the books-at-home measure and its extensions (i.e., number of children's books, number of ebooks) as well as the role of the information source (parents vs. children) for explaining students' academic language comprehension (Table 2). We additionally tested the indirect effects of parental education and HISEI on students' academic language comprehension by adding the number of books at home and the number of children's books at home as mediating variables (Table 3).

Table 2 Multiple Linear Regression Models Predicting Students' Academic Language Comprehension ($N=2353$)

Predictors	Model 1 β	Model 1 SE	Model 2 β	Model 2 SE	Model 3 β	Model 3 SE	Model 4 β	Model 4 SE	Model 5 β	Model 5 SE	Model 6 β	Model 6 SE	Model 7 β	Model 7 SE
Grade														
Grade 3[a]	.11*	.04	.12*	.04	.12*	.04	.12*	.03	.12*	.03	.13*	.03	.12*	.03
Grade 4[a]	.29*	.05	.30*	.04	.30*	.04	.30*	.04	.30*	.04	.31*	.04	.30*	.04
Language background														
Bilingual[b]	-.33*	.02	-.25*	.02	-.21*	.02	-.20*	.02	-.20*	.02	-.23*	.02	-.20*	.0
GSL[b]	-.35*	.02	-.26*	.02	-.22*	.02	.20*	.02	-.20*	.02	-.24*	.02	-.20*	.02
HISEI[c]			.28*	.03	.21*	.03	.20*	.03	.20*	.03	.25*	.03	.20*	.03
Parental education[d]			.10*	.02	.06*	.02	.06*	.02	.05*	.02	.08*	.02	.05*	.02
Source: parents — Number of books[e]					.18*	.02	.11*	.02	.11*	.03			.10*	.03
Source: parents — Number of children's books[e]							.11*	.03	.10*	.03			.10*	.03
Source: parents — Number of ebooks[f]									.01	.02				
Source: child — Number of books[e]											.06*	.03	.01	.03
Source: child — Number of children's books[e]											.07*	.02	.05	.03
R^2	.206*		.314*		.335*		.340*		.340*		.325*		.342*	
ΔR^2 (F for ΔR^2)			.108 (61.45*)		.021 (10.55*)		.005 (2.21*)		.00 (0)		.011 (4.76*)[g]		.002 (0.71)[h]	

GSL: German as a second language

[a]Reference category: Grade 2

[b]Reference category: German monolingual

[c]Highest International Socio-Economic Index of Occupational Status

[d]Parental education level (PARED) as indicated by the highest number of education years of both parents, which is based on the highest ISCED-level of both parents

[e]Assessed with a Likert scale, ranging from 1 (0–10 books) to 5 (more than 200 books)

[f]Assessed with a Likert scale, ranging from 1 (0–10 ebooks) to 5 (more than 100 ebooks)

[g]Compared to Model 2

[h]Compared to Model 4. *$p < .05$

Table 3 Estimates of Specific Indirect Effects of HISEI and Parental Education via the Number of Books on Students' Academic Language Comprehension ($N=2353$)

	Specific indirect effects	Model 4a		Model 6a	
		β	SE	β	SE
Source: parents	HISEI → Number of books →ALC	.04*	.01		
	HISEI → Number of children's books →ALC	.04*	.01		
	Parental education →Number of books→ ALC	.03*	.01		
	Parental education → Number of children's books → ALC	.03*	.01		
Source: child	HISEI → Number of books → ALC			.02*	.01
	HISEI → Number of children's books → ALC			.01*	.01
	Parental education →Number of books → ALC			.01*	.00
	Parental education → Number of children's books → ALC			.01*	.00

ALC: Academic language comprehension. HISEI: Highest International Socio-Economic Index of Occupational Status. Parental education level (PARED) as indicated by the highest number of education years of both parents, which is based on the highest ISCED-level of both parents. Number of books and children's books were assessed with Likert scales, ranging from 1 (0–10 books) to 5 (more than 200 books)

*$p < .05$

Model 1 served as a baseline model and included only the control variables grade and language background, i.e., whether students were bilingual or learned German as a second language. As expected, higher grade levels were associated with better test performance. Both bilingual students and students with German as a second language performed below their monolingual German-speaking peers. The control variables accounted for 20.6% of the variance in the outcome variable.

In Models 2 to 7, the various SES indicators were added step by step. Thus, Model 2 additionally included the families' HISEI and parental education. Both measures contributed to the explanation of students' academic language comprehension and resulted in a significant increase in the amount of explained variance ($R^2 = .314$, $\Delta R^2 = .108$). With the next three models, we aimed to investigate the importance of the number of books, children's books, and ebooks as indicated by the parents. We found positive effects for the number of books, resulting in a 2.1% increase in the amount of explained variance (Model 3). The number of children's books was a significant predictor above and beyond HISEI, parental education, and the number of books and led to a small but significant increase in explained variance (Model 4; $R^2 = .340$, $\Delta R^2 = .005$).[4] However, the number of ebooks, which we additionally entered in Model 5, did not contribute to the explanation of students' academic language comprehension. Further analyses revealed that parental HISEI

---

[4] Further analyses revealed that the amount of variance explained by parents' information on the number of books and the number of children's books was almost the same as for HISEI and parental education. However, R^2 was slightly smaller when students shared information on the number of books and children's books (Table S5 in the Supplemental Material).

and education were both significant predictors of the number of books ($\beta_{HISEI} = .39$, $p < .05$, $SE = .03$; $\beta_{parental\ education} = .24$, $p < .05$, $SE = .03$) and the number of children's books ($\beta_{HISEI} = .33$, $p < .05$, $SE = .03$; $\beta_{parental\ education} = .24$, $p < .05$, $SE = .03$). Moreover, their relation to students' academic language comprehension was mediated by the number of books and children's books, as indicated by the significant Sobel test (Sobel, 1982; Model 4a in Table 3).

To investigate the impact of the number of books and number of children's books as assessed by the children, we removed parents' information on the number of books at home and added information gathered from the students instead (Model 6). Both variables significantly contributed to the explanation of the outcome variable. Compared to Model 2, which included only the control variables as well as HISEI and parental education, additionally considering students' information on their own and their parents' books significantly increased the amount of explained variance by 1.1%, from 31.4 to 32.5%. Similar to parents' information on the number of books at home, children's information on the number of books ($\beta_{HISEI} = .26$, $p < .05$, $SE = .03$; $\beta_{parental\ education} = .14$, $p < .05$, $SE = .03$) and children's books at home ($\beta_{HISEI} = 0.18$, $p < .05$, $SE = .03$; $\beta_{parental\ education} = .11$, $p < .05$, $SE = .03$) was predicted by parental HISEI and education. Both variables mediated the relation of the two distal structural features of SES on students' academic language comprehension (Model 6a in Table 3).

Ultimately, Model 7 included all variables that were significant predictors of students' academic language comprehension in the previous models, thus investigating whether the predictive power of the number of books and the number of children's books at home differed by information source (parents vs. children). In this final model, we found that the number of books and children's books as indicated by the parents were both significant predictors of students' academic language comprehension, even when considering the controls, HISEI, and parental education. However, no effects emerged for the number of books and children's books when information was collected from the children. The amount of explained variance was 34.3%, which was almost the same as in Model 4. The results thus indicate the importance of parents' information as opposed to children's information when using the number of books at home as a predictor of academic achievement.

8 Discussion

The current study investigated the incremental validity of the books-at-home measure beyond other commonly used SES indicators (i.e., parents' occupational status and education) in explaining academic language comprehension. In particular, we examined whether selected extensions of the traditional books-at-home measure, namely, the number of children's books and the number of ebooks, increase the validity of the books-at-home measure and whether the predictive value of the number of books and the number of children's books differs by information source (i.e., whether parents or children answered the question). Additional analyses examined whether the number of books and children's books as indicated by both parents and

children mediated the relation between more distal structural SES features and students' academic language comprehension.

8.1 Predictive validity of the books-at-home measure and its extensions

We found that parents' information on the number of books and children's books at home significantly increased the amount of explained variance in students' academic language comprehension, even when considering parental education and occupational status (HISEI). The number of ebooks, however, did not contribute to the explanation of students' academic language comprehension. A similar pattern of results emerged when using children's information instead of parents' information. Thus, children's estimates, which were only assessed for the number of books and the number of their own books but not for the number of ebooks, contributed significantly to the explanation of their academic language comprehension. Moreover, parents' books and children's books, both when assessed by parents and by children, mediated the relationship between parents' occupational status and education as well as students' academic language comprehension. When simultaneously considering parents' and children's estimates of the number of books at home, however, only parents' information on the number of books and the number of children's books remained significant. The results thus show that children's information on the number of books at home is of limited predictive value compared to parents' information for explaining student achievement (see Table S5 in the Supplemental Material for additional findings in support of this interpretation). Overall, the findings indicate that when used in the parent questionnaire, the validity of the traditional books-at-home measure is not compromised by the greater availability of ebooks and can be slightly increased by additionally considering the number of children's books at home.

The present study's results corroborate previous findings that confirm the interrelatedness of different SES indicators while underlining their distinctiveness (e.g., Bukodi & Goldthorpe, 2013). Thus, although the different measures for assessing a family's SES showed substantial amounts of shared variance, parental occupation status (HISEI), parental education, number of books, and number of children's books all contributed independently to the explanation of student achievement. This suggests that they capture slightly different aspects of SES and cannot be used interchangeably. Simultaneously, the present study adds to the literature by examining the importance of the number of children's books and the number of ebooks as well as the role of the respondent (parents vs. children) in increasing the predictive value of the traditional books-at-home measure. Furthermore, mediation analyses' results confirmed and extended prior research that pointed to the mediating role of combined indices of the number of books and children's books for the relation between different SES measures and students' academic achievement (e.g., McElvany et al., 2009; McMullin et al., 2020; Myrberg & Rosén, 2009). The present findings thus further support the assumption that parents' occupational status and education should be conceived of as distal structural features of SES whose impact on students' learning outcomes can at least partly be explained by more proximal

process-oriented features such as the learning stimulation tied to books and children's books available at home (cf. Gustafsson et al., 2011; McElvany et al., 2009).

In terms of underlying mechanisms, which help to explain the relation between the number of books at home and student academic achievement, theoretical considerations suggest that the number of children's books may be a better proxy for learning stimulation and joint reading activities than the number of parents' books and, thus, may be an even more valid indicator of students' cultural capital and learning resources. However, remarkably, this measure is not typically considered in large-scale assessments (with the exception of TIMSS and PIRLS), and its individual effects in predicting student achievement (i.e., net of the number of parents' books) are usually not investigated. Hence, whereas prior research mostly used combined measures of the number of parents' and children's books and found that they were positively related to, for instance, students' reading comprehension (Gustafsson et al., 2011; McElvany et al., 2009), we established the number of children's books as an independent predictor of students' academic language comprehension.

For the number of ebooks, the present findings challenge the assumption that the process of digitalization and the greater availability of digital devices threaten the validity of the traditional books-at-home measure (cf. Schwippert, 2019). In line with prior studies on media use (Kinder-Medien-Studie, 2018; Statista, 2021) that did not report widespread use of ebooks, the vast majority of parents in our study indicated owning 0 to 10 ebooks or chose not to answer the question. We suspect that this very uneven distribution and resulting variance restriction, which differed sharply from all the other books-at-home indicators (Fig. S1 in the Supplemental Material), is the driving force for explaining the null effects in the present study (for convergent findings, see Pagel, 2016). Potential differences in the use of ebooks and printed books in joint interactions among parents and children (cf. Krcmar & Cingel, 2014; Ross et al., 2016) might additionally have come into play.

The present study is among the very few to examine parents' and children's ratings of the number of books at home (e.g., Engzell, 2018; Jerrim & Micklewright, 2014) and, to the best of our knowledge, is the first to compare the predictive value of the number of books and the number of children's books for these different information sources. In line with the studies by Engzell (2018) and Jerrim and Micklewright (2014), for instance, we found exceptionally low agreement between parents' and children's information on parents' books and children's books. Although it is not possible to determine which ratings are more accurate, we have reason to assume that elementary school children are the less reliable information source given their limited capacity for estimating amounts (Harel et al., 2007). Additionally, considering the stronger relations between parents' estimates and students' academic achievement compared to students' estimates, our findings support the inclusion of the books-at-home measure in the parent questionnaire rather than in the student questionnaire (see Hovestadt & Schneider, 2021, for convergent findings regarding parental education).

8.2 Limitations, future directions, and conclusion

Several limitations exist in the current study. First, no information was available on the processes that occur within families, which might help to explain the effects of the number of books and the number of children's books on students' learning outcomes. While it can reasonably be assumed that books at home form an important basis for home literacy activities, such as joint reading activities or talking about reading experiences (e.g., Martin & Mullis, 2013; McElvany et al., 2009), future studies should deliberately assess students' home literacy activities to inform our understanding of the role of parents' books compared to children's books for student outcomes.

Second, the present analyses are based on cross-sectional data and, thus, do not allow for drawing causal inferences. We assumed and modeled the various SES indicators as predictors of students' academic language comprehension and can reasonably exclude reverse causality for most relations. Parents' number of books and, in particular, their occupational status and education are most likely unaffected by their children's academic language comprehension. However, reciprocal relations may occur between academic language comprehension and children's own books, as children with greater mastery of the academic register may demand and be supplied with more books than students who are less proficient in academic language comprehension (cf. Mol & Bus, 2011).

Third, students' academic language comprehension was the only outcome measure involved in our analyses. Academic language proficiency has been shown to be substantially related to competencies in a variety of domains, such as reading comprehension, mathematics, and science (e.g., Schuth et al., 2017; Volodina et al., 2021b), thus confirming it as a meaningful variable for investigating social inequalities in student achievement. However, the predictive value of the books-at-home measure and its extensions might be smaller for less language-bound measures.

Fourth, although we controlled for parents' occupational status and education, which are both substantially related to the various books-at-home measures and students' academic language comprehension, further measures that might help capture an even more nuanced picture of the relation between the number of books at home and student achievement were not considered. Specifically, information on parents' home ownership, living space, and recent or upcoming moves, which all may relate to the number of books at home were not included in the dataset. In particular, frequent relocations may pair with a diminished personal book stock independent of a person's occupational status and education.

Despite these limitations and open questions, which are subject to future research, the present study's results may serve as an important basis for selecting and assessing SES indicators in social research. In particular, they increase our knowledge of the validity of the books-at-home measure, which is ubiquitously used in surveys and large-scale assessments but the quality of which has only rarely been scrutinized.

Declarations

Conflict of interest We have no known conflict of interest to disclose.

Ethical approval In line with the German regulations for conducting empirical studies in schools, our study was evaluated and approved by the ministries of education of the six participating federal states. This procedure covered scrutiny of all materials used in the study, such as cover letters to schools, teachers, students, and parents (including information on the voluntary nature of participation, etc.), survey materials (questionnaires, tests, items, etc.), and procedures (e.g., for data protection and anonymization). In addition, we received project approval by the Ethics Committee of the University of Bamberg. Informed consent was obtained from all study participants.

References

Achugar, M., & Schleppegrell, M. J. (2005). Beyond connectors: The construction of cause in history textbooks. *Linguistics and Education, 16*(3), 298–318. https://doi.org/10.1016/j.linged.2006.02.003

Bailey, A. L. (Ed.). (2007). *The language demands of school. Putting academic English to the test*. Yale University Press.

Baydar, N., & Akcinar, B. (2015). Ramifications of socioeconomic differences for three year old children and their families in Turkey. *Early Childhood Research Quarterly, 33*(4), 33–48. https://doi.org/10.1016/j.ecresq.2015.05.002

Berendes, K., Vajjala, S., Meurers, D., Bryant, D., Wagner, W., Chinkina, M., & Trautwein, U. (2018). Reading demands in secondary school: Does the linguistic complexity of textbooks increase with grade level and the academic orientation of the school track? *Journal of Educational Psychology, 110*(4), 518–543. https://doi.org/10.1037/edu0000225

Bourdieu, P. (1986). The forms of capital. In J. E. Richardson (Ed.), *Handbook of theory of research for the sociology of education* (pp. 241–258). Greenword Press.

Bukodi, E., & Goldthorpe, J. H. (2013). Decomposing 'social origins': The effects of parents' class, status, and education on the educational attainment of their children. *European Sociological Review, 29*(5), 1023–1039. https://doi.org/10.1093/esr/jcs079

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

Carbonaro, W. J. (1998). A little help from my friend's parents: Intergenerational closure and educational outcomes. *Sociology of Education, 71*(4), 295–313. https://doi.org/10.2307/2673172

Chatterjee, S., & Simonoff, J. S. (2013). *Handbook of regression analysis*. John Wiley & Sons. https://doi.org/10.5860/choice.51-0931

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*, 95–120.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330–351. https://doi.org/10.1037/1082-989x.6.4.330

Conger, R. D., & Donnellan, M. B. (2007). An interactionist perspective on the socioeconomic context of human development. *Annual Review of Psychology, 58*, 175–199. https://doi.org/10.1146/annurev.psych.58.110405.085551

Engzell, P. (2018). What do books in the home proxy for? A cautionary tale. *Sociological Methods & Research, 50*(4), 1487–1514. https://doi.org/10.1177/0049124119826143

Engzell, P., & Jonsson, J. O. (2015). Estimating social and ethnic inequality in school surveys: Biases from child misreporting and parent nonresponse. *European Sociological Review, 31*(3), 312–325. https://doi.org/10.1093/esr/jcv005

Erikson, R., & Jonsson, J. O. (1996). Explaining class inequality in education: The Swedish test case. In R. Erikson & J. O. Jonsson (Eds.), *Can education be equalized? The Swedish case in comparative perspective* (pp. 1–64). Westview Press.

Eriksson, K., Lindvall, J., Helenius, O., & Ryve, A. (2021). Socioeconomic status as a multidimensional predictor of student achievement in 77 societies. *Frontiers in Education*. https://doi.org/10.3389/feduc.2021.731634

Erola, J., Jalonen, S., & Lehti, H. (2016). Parental education, class and income over early life course and children's achievement. *Research in Social Stratification and Mobility, 44*, 33–43. https://doi.org/10.1016/j.rssm.2016.01.003

Evans, M. D. R., Kelley, J., Sikora, J., & Treiman, D. J. (2010). Family scholarly culture and educational success: Books and schooling in 27 nations. *Research in Social Stratification and Mobility, 28*(2), 171–197. https://doi.org/10.1016/j.rssm.2010.01.002

Fitzgerald, J., Elmore, J., Relyea, J. E., & Stenner, A. J. (2020). Domain-specific academic vocabulary network development in elementary grades core disciplinary textbooks. *Journal of Educational Psychology, 112*(5), 855–879. https://doi.org/10.1037/edu0000386

Fitzgerald, J., Relyea, J. E., & Elmore, J. (2022). Academic vocabulary volume in elementary grades disciplinary textbooks. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000735

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

Grolig, L., Cohrdes, C., Tiffin-Richards, S. P., & Schroeder, S. (2019). Effects of preschoolers' storybook exposure and literacy environments on lower level and higher level language skills. *Reading and Writing, 32*(4), 1061–1084. https://doi.org/10.1007/s11145-018-9901-2

Gustafsson, J. E., Hansen, K. Y., & Rosén, M. (2011). Effects of home background on student achievement in reading, mathematics, and science at the fourth grade. *TIMSS and PIRLS Relationship Report, 4*, 181–287.

Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson Prentice Hall.

Hällsten, M., & Thaning, M. (2021). Wealth as one of the "Big Four" SES dimensions in intergenerational transmissions. *Social Forces, 100*(4), 1533–1560. https://doi.org/10.1093/sf/soab080

Hanushek, E. A., & Woessmann, L. (2011). The economics of international differences in educational achievement. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of economics of education* (pp. 89–200). Elsevier.

Harel, B. T., Cillessen, A. H. N., Fein, D. A., Bullard, S. E., & Aviv, A. (2007). It takes nine days to iron a shirt: The development of cognitive estimation skills in school age children. *Child Neuropsychology, 13*(4), 309–318. https://doi.org/10.1080/09297040600837354

Harlaar, N., Dale, P. S., & Plomin, R. (2007). Reading exposure: A (largely) environmental risk factor with environmentally-mediated effects on reading performance in the primary school years. *The Journal of Child Psychology and Psychiatry, 48*(12), 1192–1199. https://doi.org/10.1111/j.1469-7610.2007.01798.x

Harwell, M., Maeda, Y., Bishop, K., & Xie, A. (2017). The surprisingly modest relationship between SES and Educational Achievement. *The Journal of Experimental Education, 85*(2), 197–214. https://doi.org/10.1080/00220973.2015.1123668

Heppt, B., & Stanat, P. (2020). Development of academic language comprehension of German monolinguals and dual language learners. *Contemporary Educational Psychology, 62*, 101868. https://doi.org/10.1016/j.cedpsych.2020.101868

Heppt, B., Köhne-Fuetterer, J., Eglinsky, J., Volodina, A., Stanat, P., & Weinert, S. (2020). *BiSpra 2–4. Test zur Erfassung bildungssprachlicher Kompetenzen bei Grundschulkindern der Jahrgangsstufen 2 bis 4* [BiSpra 2–4. Test for the assessment of academic language skills for elementary school children from grades 2 through 4]. Waxmann.

Hovestadt, T., & Schneider, T. (2021). Liefern Jugendliche valide Informationen zum Bildungsstand ihrer Eltern in standardisierten Erhebungen? Befunde zu Schülerinnen und Schülern der 9. Jahrgangsstufe in Deutschland [Can adolescents provide valid information on parents' educational attainment in standardized surveys? Results from a survey conducted in Grade 9 in Germany]. *Zeitschrift für Erziehungswissenschaft, 24*(3), 715–742. https://doi.org/10.1007/s11618-021-01016-5

International Labour Office. (2012). *International standard classification of occupations: Structure, group definitions and correspondence tables*, vol. 1. International Labour Organization.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with application in R*. Springer.

Jerrim, J., & Micklewright, J. (2014). Socio-economic gradients in children's cognitive skills: Are cross-country comparisons robust to who reports family background? *European Sociological Review, 30*(6), 766–781. https://doi.org/10.1093/esr/jcu072

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity, 52*(4), 1957–1976. https://doi.org/10.1007/s11135-017-0584-6

Kinder-Medien-Studie. (2018). https://kinder-medien-studie.de/wp-content/uploads/2018/08/KMS2018_Berichtsband_v2.pdf.

Klein, D. (2011). MIVIF: Stata module to calculate variance inflation factors after mi estimate regress. *Statistical Software Components*, S457254. https://ideas.repec.org/c/boc/bocode/s457254.html.

Krcmar, M., & Cingel, D. P. (2014). Parent–child joint reading in traditional and electronic formats. *Media Psychology, 17*(3), 262–281. https://doi.org/10.1080/15213269.2013.840243

Leszczensky, L., & Wolbring, T. (2022). How to deal with reverse causality using panel data? Recommendations for researchers based on a simulation study. *Sociological Methods & Research, 51*(2). 837–865. https://doi.org/10.1177/0049124119882473

Martin, M. O., & Mullis, I. V. S. (2013). *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade—Implications for early learning*. TIMSS & PIRLS International Study Center.

McElvany, N., Becker, M., & Lüdtke, O. (2009). Die Bedeutung familiärer Merkmale für Lesekompetenz, Wortschatz, Lesemotivation und Leseverhalten [The meaning of family characteristics for reading proficiency, vocabulary, reading motivation and reading behavior]. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie, 41*(3), 121–131. https://doi.org/10.1026/0049-8637.41.3.121

McMullin, P., McGinnity, F., Murray, A., & Russell, H. (2020). What you do versus who you are: Home-learning activities, social origin and cognitive skills among young children in Ireland. *European Sociological Review, 36*(4), 610–625. https://doi.org/10.1093/esr/jcaa012

Meneses, A., Uccelli, P., Santelices, M. V., Ruiz, M., Acevedo, D., & Figueroa, J. (2018). Academic language as a predictor of reading comprehension in monolingual Spanish-speaking readers: Evidence from Chilean early adolescents. *Reading Research Quarterly, 53*(2), 223–247. https://doi.org/10.1002/rrq.192

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137*(2), 267–296. https://doi.org/10.1037/a0021890

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012a). *TIMSS 2011 international results in mathematics*. TIMSS & PIRLS International Study Center, International Association for the Evaluation of Educational Achievement.

Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012b). *PIRLS 2011 international results in reading*. TIMSS & PIRLS International Study Center, International Association for the Evaluation of Educational Achievement.

Muthén, L. K., & Muthén, B. O. (1998–2019). *Mplus user's guide*, 7th edn. Muthén & Muthén.

Myrberg, E., & Rosén, M. (2009). Direct and indirect effects of parents' education on reading achievement among third graders in Sweden. *British Journal of Educational Psychology, 79*(4), 695–711. https://doi.org/10.1348/000709909X453031

OECD. (1999). *Classifying educational programmes: Manual for ISCED-97 implementation in OECD countries*. OECD Publishing. http://www.oecd.org/edu/skills-beyond-school/1962350.pdf.

OECD. (2009). Appendices. In OECD (Ed.), *PISA 2006 Technical Report*. OECD Publishing. https://doi.org/10.1787/9789264048096-en.

OECD. (2014). *PISA 2012. Technical Report*. OECD Publishing.

OECD. (2019). *PISA 2018 Results (Volume II): Where all students can succeed*. OECD Publishing. https://doi.org/10.1787/b5fd1b8f-en.

Pagel, L. (2016). *Die Bücherfrage als Indikator des kulturellen Kapitals in der Bildungsforschung. Lässt sich die Validität durch eine differenziertere Erfassung erhöhen? [The book question as an indicator of cultural capital in educational research. Can validity be increased by a more differentiated recording?]* Unpublished master's thesis, Humboldt-Universität zu Berlin.

Park, H. (2008). Home literacy environments and children's reading performance: A comparative study of 25 countries. *Educational Research and Evaluation, 14*(6), 489–505. https://doi.org/10.1080/13803610802576734

Richter, D., Böhme, K., Bastian-Wurzel, J., Pant, H. A., & Stanat, P. (2014). *IQB-Ländervergleich 2011. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente* [The IQB national assessment study 2011. Scale manual of the assessment instruments], vol. 6. Humboldt-Universität zu Berlin, IQB. https://doi.org/10.18452/3127.

Rideout, V. (2014). *Learning at home: families' educational media use in America*. The Joan Ganz Cooney Center at Sesame Workshop.

Robitzsch, A., Kiefer, T., & Wu, M. (2021). TAM: Test Analysis Modules. https://CRAN.R-project.org/package=TAM.

Rodgers, D. (1974). Which connectives? Signals to enhance comprehension. *Journal of Reading, 17*(6), 462–466.

Ross, K. M., Pye, R. E., & Randell, J. (2016). Reading touch screen storybooks with mothers negatively affects 7-year-old readers' comprehension but enriches emotional engagement. *Frontiers in Psychology, 7*, 1728. https://doi.org/10.3389/fpsyg.2016.01728

Schleppegrell, M. J. (2004). *The language of schooling*. Lawrence Erlbaum. https://doi.org/10.4324/9781410610317

Schuth, E., Köhne, J., & Weinert, S. (2017). The influence of academic vocabulary knowledge on school performance. *Learning and Instruction, 49*, 157–165. https://doi.org/10.1016/j.learninstruc.2017.01.005

Schwippert, K. (2019). Was wird aus den Büchern? Sozialer Hintergrund von Lernenden und Bildungsungleichheit aus Sicht der international vergleichenden Erziehungswissenschaft [What's about the books? Social background of students and educational opportunities from the perspective of international large-scale surveys]. *Journal for Educational Research Online, 11*(1), 92–117. https://doi.org/10.25656/01:16789

Sieben, S., & Lechner, C. M. (2019). Measuring cultural capital through the number of books in the household. *Measurement Instruments for the Social Sciences, 2*(1), 1–6. https://doi.org/10.1186/s42409-018-0006-0

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science, 328*(5977), 450–452. https://doi.org/10.1126/science.1182597

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In E. Leinhardt (Ed.), *Sociological Methodology* (pp. 290–312). American Sociological Association. https://doi.org/10.2307/270723.

Statista. (2021). *Statista-Dossier zum E-Book-Markt in Deutschland* [Statista dossier on the ebook market in Germany]. https://de.statista.com/statistik/studie/id/6689/dokument/e-books-statista-dossier/

Uccelli, P., Demir-Lira, O. E., Rowe, M. L., Levine, S., & Goldin-Meadow, S. (2019). Children's early decontextualized talk predicts academic language proficiency in midadolescence. *Child Development, 90*(5), 1650–1663. https://doi.org/10.1111/cdev.13034

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

Volodina, A., Weinert, S., & Mursin, K. (2020). Development of academic vocabulary across primary school age: Differential growth and influential factors for German monolinguals and language minority learners. *Developmental Psychology, 56*(5), 922–936. https://doi.org/10.1037/dev0000910

Volodina, A., Heppt, B., & Weinert, S. (2021a). Effects of socioeconomic status and language use on academic language proficiency in children with a migration background: An evaluation using quantile regressions. *Contemporary Educational Psychology*, *65,* 101973. https://doi.org/10.1016/j.cedpsych.2021.101973

Volodina, A., Heppt, B., & Weinert, S. (2021b). Relations between the comprehension of connectives and school performance in primary school. *Learning and Instruction, 74,* 101430. https://doi.org/10.1016/j.learninstruc.2020.101430

Warm, T. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450. https://doi.org/10.1007/BF02294627

Weinert, S., Stanat, P., Heppt, B., Volodina, A., & Eglinsky, J. (2020). *BiSpra-Aufgaben: Weiterentwicklung zu einem diagnostisch nutzbaren Testinstrument und Prüfung der Sensitivität für Fördereffekte (Verbundprojekt). Schlussbericht 2019.* [BiSpra-Transfer: Development and validation of a standardized test instrument (joint project). Final report 2019]. University of Bamberg. https://doi.org/10.18452/22352.

Birgit Heppt holds an interim professorship for educational psychology at the Institute for Educational Sciences at the Humboldt-Universität zu Berlin, Germany. She is a fellow at the College for Interdisciplinary Educational Research (CIDER) and received her PhD in psychology from Humboldt-Universität zu Berlin, Germany. Her research focuses on the assessment and development of academic language proficiency, teachers' professional development for language support, and immigration-related inequalities in student achievement.

Melanie Olczyk is a research fellow (postdoc) for the project "Educational integration of refugee children and youth in Germany (EDIREG)", University of Halle. Moreover, she is a fellow at the College for Interdisciplinary Educational Research (CIDER). She received her PhD in sociology at the University of Bamberg, Germany. Her research focuses on social and migrant-specific educational inequality, social networks and social capital, and international comparisons.

Anna Volodina is a postdoctoral researcher at the University of Bamberg, Germany. She received her PhD in psychology from the University of Kiel, Germany. Her research focuses on the assessment and development of (academic) language proficiency, school success, school transitions, and socioeconomic inequalities in child development.

Authors and Affiliations

Birgit Heppt[1] [ID] · **Melanie Olczyk[2]** [ID] · **Anna Volodina[3]** [ID]

Melanie Olczyk
melanie.olczyk@soziologie.uni-halle.de

Anna Volodina
anna.volodina@uni-bamberg.de

[1]   Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

[2]   Institute of Sociology, Martin-Luther-Universität Halle-Wittenberg, Paracelsusstraße 22, 06114 Halle (Saale), Germany

[3]   University of Bamberg, Markusplatz 3, 96047 Bamberg, Germany