Bittmann, Felix

Reading Begets Reading? : Disentangling the Dynamic Interplay Between Reading Competence and Reading Exposure with a Special Focus on Gender Differences

Date of secondary publication: 14.03.2025

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-1070201

Primary publication

Bittmann, Felix (2024): Reading Begets Reading? : Disentangling the Dynamic Interplay Between Reading Competence and Reading Exposure with a Special Focus on Gender Differences, in: Reading research quarterly : RRQ, Hoboken, NJ: Wiley-Blackwell, Vol. 60, Nr. 1, e566, pp. 1–16, doi: 10.1002/rrq.566.

INTERNATIONAL LITERACY ASSOCIATION
READING RESEARCH QUARTERLY

Reading Begets Reading? Disentangling the Dynamic Interplay Between Reading Competence and Reading Exposure with a Special Focus on Gender Differences

Felix Bittmann
Leibniz Institute for Educational Trajectories, Bamberg, Germany

ABSTRACT

Do children with better reading competence read more, or do avid readers increase their reading competence? This highly relevant question has been discussed for many years, yet conclusive results are rare. Previous studies suffer from small sample sizes or omitted variable bias, rendering their findings questionable. We provide new insight using large-scale German panel data ($N > 5100$). By surveying secondary school students (initial age = 11.1 years, SD = .50 years) in grades 5, 7, and 9, we can trace their reading competence and exposure over time. We estimate random intercept cross-lagged panel models (RI-CLPMs) to analyze time-dependent within-student changes and dynamic feedback loops. We include relevant control variables to account for potential confounding and conduct robustness checks using a classical CLPM. Our findings show no evidence that reading more results in better reading competence. However, higher reading competence may lead to increased reading exposure. Testing for gender differences reveals that these conclusions hold for both boys and girls and that no interactions with gender are detectable, despite girls tending to read more and having a higher reading competence on average.

Introduction

Reading is one of the most fundamental cultural techniques humans have developed in their history and a required skill for virtually everyone living in modern societies. However, the variability in reading skill and competence is large, and some individuals read better than others. One of the central questions in reading psychology is whether these differences can be explained, to some extent, by the amount individuals read, which we refer to as reading exposure (Cunningham & Stanovich, 1990). It is well known that muscles that work harder grow more; does the same principle apply to reading competence? Or is the opposite true, meaning that highly-skilled readers read more because it is easier for them and they enjoy it more? Disentangling this dynamic feedback process between reading competence and exposure is a question we want to answer. To be able to give practically relevant advice, we focus on the approximation of causal effects. In addition, we would like to study gender differences in particular. It is well known that reading ability in students differs by gender, and girls usually have higher reading competence than boys (Lietz, 2006). The question arises whether this advantage concerning competence influences the interplay between exposure and competence. For a convenient introduction to the topic, the first chapter is structured

INTERNATIONAL LITERACY ASSOCIATION

Reading Research Quarterly, 2025
60:e566 | doi:10.1002/rrq.566

as follows: First, the correlational relationship between reading exposure and competence is empirically established and theoretically explained. Afterwards, the literature on the topic, with a special focus on causal models, is recapped. Lastly, the question of what makes a good causal analysis is discussed and the research goals of the paper summarized.

It is well-researched that avid readers are also better readers (Anderson et al., 1988; Cunningham & Stanovich, 1997). A meta-analysis of almost 100 studies demonstrates the correlation between competence and exposure (Mol & Bus, 2011). As the included studies stem from various cultural contexts and populations and are using different measurements and methods, we can be confident that the association or correlation between reading exposure and competence is firmly established. However, does this imply any causal relationship? While correlations can be helpful and are usually a first analysis step, the benefits for policy advice and instructional planning are limited as a simple correlation does often not imply causation. It would be highly relevant to determine whether there is a causal relationship between the two constructs as only this knowledge should be used to advise stakeholders. Consequently, this is the first research question: is there a causal relationship between reading exposure and reading competence and in which direction does it go?

Regarding this are three potential explanations. First, the relation is bidirectional and competence and reading influence each positively other in a virtuous cycle, also known as the Matthew effect of reading (Stanovich, 2009). The second possibility is that exposure only affects competence positively. This assumption is supported by the Simple View of Reading model (SVR) (Hoover & Gough, 1990). SVR assumes that reading competence depends on (a) decoding and (b) linguistic comprehension. A more frequent exposure to words helps children acquire orthographic knowledge, which facilitates the decoding process (Share, 1999). If this process becomes faster and easier, more cognitive resources are available for comprehension (Mol & Bus, 2011). This model leads to the prediction that a more frequent exposure leads to higher reading competences. The third option is that better readers are also more avid readers, because reading is easy for them and they enjoy it more as it requires little cognitive resources. Since the process of reading (decoding and comprehension) happens quickly and automatically, they can focus on the content they are reading, which makes reading a fun activity that is taken up often. The SVR is also applicable in that case as it explains how cognitive burden is reduced for the better readers due to their previous experience and competence (Juel, 1988). Consequently, we argue that the SVR is a good theoretical starting point to understand how reading exposure and competence are related to each other. Based on these first theoretical outlook, what is the current state of research?

A literature review reveals that the overall amount of empirical evidence regarding the co-development of reading competence and exposure could be more extensive and convincing. One study with longitudinal data ($N<400$) with primary school students reports that reading exposure develops mostly independently of competence (Aarnoutse & Van Leeuwe, 1998). A study with students in first and second primary school grades ($N<200$) does find evidence for crossed effects, that is, statistically significant co-development of key constructs (Leppänen et al., 2005). A third study, including 436 twin pairs, reports statistically significant effects from achievement on exposure. However, the other crossed path showed no such effect (Harlaar et al., 2011). Two more studies report similar effects from competence to exposure (Becker et al., 2010; Cunningham & Stanovich, 1997). One study with German data ($N<800$) reports that there is no influence of reading amount in grade 4 on reading literacy in grade 6 when reading motivation and earlier reading literacy are controlled for (Becker et al., 2010). A further study with about 2500 students investigates the co-development from school grades one to nine and reports that from grade four onward, both crossed paths are statistically significant (Torppa et al., 2020). Two other studies take a different approach and use twin-data to model causality cross-sectionally and conclude with large samples that reading skill has a causal effect on exposure but there is no effect of the reversed pathway (Erbeli et al., 2020; Van Bergen et al., 2018). Another study with Finnish longitudinal data ($N=200$) computes multiple models (RI-CLPMs and CLPMs) and finds evidence for a reciprocal relationship between exposure and reading fluency (Van Bergen et al., 2021). A meta-study concludes that reading exposure does influence reading achievement (Allington & McGill-Franzen, 2021). In contrast to this finding, another recent meta-study investigates *digital* reading habits and concludes that "…leisure digital reading habits do not pay off as much as traditional print reading" (Altamura et al., 2023, p. 25). To summarize, the evidence is mixed. The pathway from reading competence to reading exposure appears more consistently than the reversed pathway. Regarding this other pathway, from exposure to competence, the kind of reading material also makes a difference. It is doubtful whether a higher reading exposure always results in better competence.

To add to the initial causal question, we would like to discuss how gender can also significantly influence the interplay of reading competence and exposure, which has rarely been done in previous research. This disregard is surprising given that it is a well-established fact that girls outperform boys consistently in terms of reading achievement. A large analysis of more than one million students reports that girls exceed boys in all OECD countries regarding reading performance (Van Hek et al., 2019). A meta-study including more than 40 studies concludes that girls are also consistently more avid readers, as they read more and more diverse genres, even if, on average, they prefer fictional texts compared to boys, who prefer non-fiction texts (Jabbar & Warraich, 2023). These differences indicate that teaching might also be different for the genders as boys prefer other genres than girls. This means, access to enough reading material is not enough, as gender-specific reading material is beneficial (Millard, 1998, pp. 49–51). This difference in reading preference also has consequences for adults, as women understand fictional texts better while men have an advantage with non-fiction texts (Thums et al., 2021). Given these stable differences in gender preferences, exposure, and competence, it seems highly relevant to test whether the dynamic co-development of reading exposure and competence differs by gender. Potentially, one could suspect that girls as more avid readers reach a higher proficiency earlier and reading even more in the higher school grades adds little to their competence development as they have already recognized and decoded the majority of the relevant words, according to the Simple View of Reading. If girls have already out-maxed their decoding skills at an earlier age, more reading exposure might not benefit them anymore. This could be a contrast to boys, who need more time to make the same progress since they read less, which means that they have a disadvantage in respect to the cumulative reading amount in their development. However, it is still possible that the general process is identical for boys and girls, even if girls perform on a higher level concerning competence and exposure. If gender differences are found, this could suggest different teaching strategies, as they have already been found regarding attention differences between boys and girls (Logan & Johnston, 2010, p. 179). This is the second research question: is the relationship between reading exposure and reading competence similar for boys and girls or not?

Lastly, we believe it is relevant to discuss the aspects of causality, why most studies do not investigate them, and what is required for a causal framework. Many of the previously cited studies only report associations, as the requirements on the data are lower. Causal frameworks, in contrast, usually either require experiments (which are expensive and often difficult to justify ethically when it comes to education), longitudinal data (which can take many years to collect) or genetic information (which are also expensive or require rather special populations, such as twins). To be precise, what exactly is required to give a definite answer when one is explicitly interested in a causal analysis (Hernán, 2018; Morgan & Winship, 2015)? We define four requirements that are paramount to answering questions about causality in the case of reading (where alternatives, such as experiments, are limited). First, longitudinal data are highly beneficial (Brüderl & Ludwig, 2015; Rohrer & Murayama, 2023). Suppose there is only a single measurement (resulting in cross-sectional data). In that case, only correlations or associations can be deduced, but the direction of causality can only be theoretically assumed but never proven. The separation of cause and effect is usually not feasible when two variables of interest strongly depend on each other and co-develop dynamically over time. Exceptions are studies using genetic or twin-data, where cross-sectional data can be sufficient to approximate causal effects, yet these types of data come with other challenges, such as limited generalization (McGue et al., 2010; Morley, 2005). Second, relevant confounders need to be accounted for (Morgan & Winship, 2015; Rohrer & Murayama, 2023). Even with longitudinal data, if both the cause and the effect depend on a third, exogenous variable, the apparent causal relationship can vanish as soon as the confounder is considered. Panel data with multiple measurement points do not automatically correct for this pitfall, and theoretically guided controls are often required, even in state-of-the-art statistical models (Lüdtke & Robitzsch, 2021). Third, the statistical approach needs to be adequate. In the past, the classical cross-lagged panel (CLPM) model has been a popular choice. However, newer studies outline that this model is sometimes inadequate to approximate causal effects, and derived models, such as the random intercept cross-lagged panel model, are better suited, depending on the concrete research question (Lucas, 2023). Fourth and final, the sample size needs to be large enough. Many studies suffer from small N, which results in low power and unstable findings (Marszalek et al., 2011). Small effects are especially difficult to detect and small case numbers can result in incorrect conclusions. Small samples also make it difficult to have a high external validity as sample characteristics can deviate from the overall population due to random sampling error, even when probability sampling is carried out correctly (Tipton et al., 2017). In these cases, it remains doubtful whether the results of the sample also apply to the general population of interest, which can be a severe limitation.

To summarize our research goals, the present study will provide new and high-quality evidence. We can present robust and high-powered models using a large-scale panel dataset of German secondary school students. As the data stem from schools all over Germany, they are also much more representative of the general student population than many previous studies, which often had to rely on a few sampled schools, which affects external validity. Our models include many potential confounders and use state-of-the-art statistical modeling. By doing so, we can approximate causal effects and shed new light on general reading processes and potential gender differences. For the first time, we can test whether the dynamic interplay between reading exposure and reading competence is similar for boys and girls or not. By doing so we can go beyond what previous studies have reported to far, especially in regards of the stability of temporal modeling with a focus on causality. The potentially most relevant study in the context of temporal co-development of reading exposure and competence is limited through its rather small sample size and the omission of relevant control variables (Van Bergen et al., 2021). Our analyses will fill that research gap. We believe that answering these questions contributes to the current research literature and provides valuable and practical insights for parents, teachers, policy-makers, and other stakeholders.

Method

Data and Participants

We utilized German National Educational Panel Study (NEPS) data to answer posed research questions[1]. The NEPS is the most ambitious research project to study the relevance of education throughout the life course. To be concrete, starting cohort (SC) 3 was utilized, which initially sampled students at the very beginning of secondary education (right after the transition from primary schooling when students are approximately 10 to 11 years old) (NEPS Network, 2024). A representative sample of secondary schools was drawn, and two classrooms in grade 5 were randomly selected within each school. All students in these classrooms were invited to participate in the NEPS. Comprehensive achievement tests for reading competence were conducted in grades 5, 7, and 9 (corresponding to survey waves 1, 3, and 5). The NEPS provides suitable data since not only student characteristics (such as age, gender, reading exposure, and reading competence) are included but also a wide range of relevant background information on the social origin and family of the student by also inviting the parents to participate in the surveys. This enabled us to minimize the potentially biasing influence of confounders. The initial sampling frame in grade 5 consisted of 5778 students who participated in the first survey. We restricted our sample slightly to meet our needs. We excluded a large refreshment sample drawn in grade 7 since, for these students, no reading competence is available for grade 5. Our analyses only included students who were part of NEPS SC3 since the start of grade 5. We also excluded students who did not participate in the first reading competence tests in grade 5, which reduced the sample size to 5193. This step also excluded all students who attended a special needs school (*Förderschule*) in grade 5 since no competence tests were administered for this special group. Information for students who dropped out later or have other kinds of missing data (item-nonresponse) is imputed, described in more detail below. The total number of schools in the analytical sample is 234. Student data were collected from September 2010 to January 2011 (wave 1), October 2012 to June 2013 (wave 3), and from October 2014 to January 2015 (wave 5).

Measures

Reading Competence

Reading competence was measured with paper-based achievement tests developed by the NEPS. They were designed to be comparable to established instruments and surveys, such as PISA, and have a similar testing framework and rationale. They are not curricular (focused on school subjects) but rather see competence as an ability to function successfully in society (Weinert et al., 2019). The designers of the testing framework outline: "Research focuses on the abilities to read a text and understand it appropriately –both as a whole and in its single statements. The emphasis is on understanding what is in the text and not primarily on memory performance for text material that has been read but is no longer available." (Gehrer et al., 2013, p. 60). This kind of test-design was well-suited for our research goals as reading exposure (discussed in detail below) focuses on typical and daily reading activities of students, which might have an impact on how well students read and understand texts in daily and typical situations (and not only in special test-situations). Each reading test has been carefully developed: "…the instruments developed in NEPS have to go through several phases of cognitive interviews […], smaller and larger preliminary studies, and large pilots (feasibility studies). Basically, the pool of texts and accompanying items developed for each starting cohort is four times as large as the final selection. […] The complete pool of test material is piloted on the target population." (Gehrer et al., 2013, p. 65).

A comparable scaling procedure was chosen (Pohl & Carstensen, 2013) to generate unidimensional proficiency scores. The tests were linked to a common scale to enable mean-level comparisons between grades and account for increasing competence over time (Fischer et al., 2019). The score was measured using 32 items in grade 5, 40 items in grade 7, and 46 items in grade 9[2]. The items utilized different response formats, such as multiple-choice or matching tasks. A branched design has been implemented so that different booklets are assigned to students of different abilities following the first survey wave (Pohl, 2013). All tests followed a common construction framework (Gehrer et al., 2013) and used five text types (i.e., information, instruction, advertising, commenting, and literary texts). The items were chosen to measure three different cognitive requirements (finding information in the text, making text-related conclusions, reflecting, and assessing). Memory effects were prevented since items were wave-specific and never repeated. An anchor-group design was employed to enable linking across measurement occasions that relied on independent link samples (Fischer et al., 2016). Multiple linking methods have been tested empirically to select the method with the highest statistical quality. This demonstrated that unidimensionality and correct differential item functioning were retained (Fischer et al., 2019). The reliabilities of the reading competence scores are fine (.77 in grade 5; .83 in grade 7; .81 in grade 9) (Gnambs & Lockl, 2023). The reading competence scores were *z*-standardized by grade 5; the mean was fixed to zero in this survey wave, and the standard deviation was one.

Reading Exposure

Regarding the type of reading material most relevant, many previous studies focused on reading for pleasure rather than reading for school since these associations are usually stronger and become even stronger as students age (Erbeli et al., 2020; Van Bergen et al., 2018). The association between reading competence and exposure is stronger when printed material such as books is investigated and not digital or online reading (Torppa et al., 2020). We followed these research strands and focused mostly on printed reading for-pleasure material outside school to measure reading exposure, for which we use the grades' 5, 7, and 9 student questionnaires. The following 12 items were available in each:

- a/b: "How much time do you usually spend reading outside of school? Please consider all possible opportunities you have for reading, in other words not only books or magazines, but also e-mails or the internet. On a normal school day / On a normal non-school day." The response scale was a Likert-type with values ranging from 1 "not at all" to 5 "more than 2 hours".
- c/g: "How often do you normally read in your spare time: … detective novels, thrillers, horror or fantasy books, such as Harry Potter or Lord of the Rings? / … classics of children's youth literature by authors, such as Erich Kästner or Otfried Preußler? / … nonfiction books? / … comic books? / … other?"
- h/l: "Do you read the following newspaper or magazines? Local newspaper(s) / Tabloids, such as *BILD*, *BZ* / Children's and youth pages in other supraregional papers, such as *Süddeutsche Zeitung* (*SZ*) or *Frankfurter Allgemeine* (*FAZ*) / Magazines, such as *Dein SPIEGEL*, *FOCUS Schule* or *GEOlino* / Other magazines for younger readers such as *Tierfreund*, *hey!*, *Bravo Sport* or *Popcorn*".

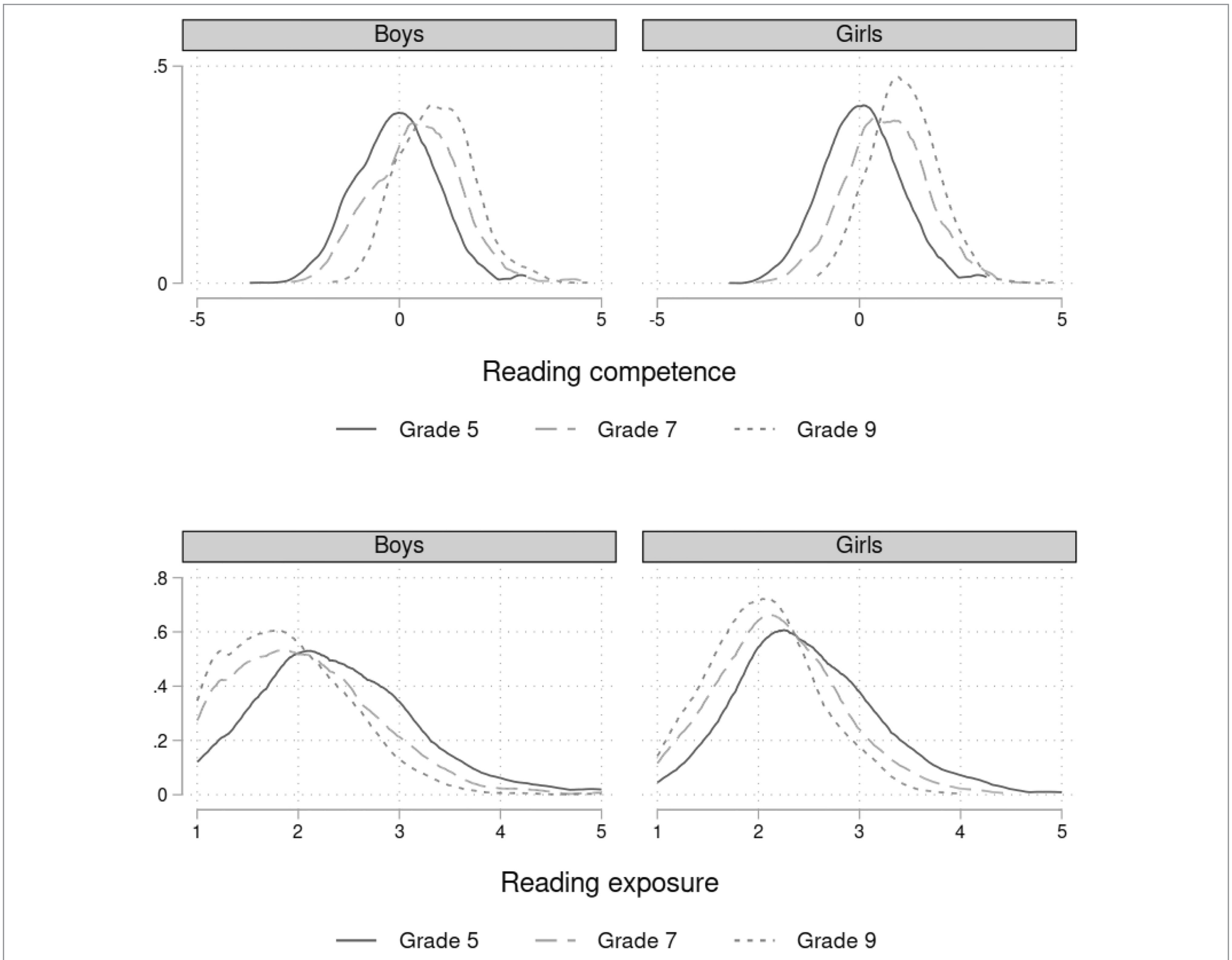
Note that these items were an adaption from the 2000 PISA study (Kunter et al., 2002) and the 2006 KOALA-S study (Ditton, 2007). The response scale for items c to l was also Likert-type with values ranging from 1 "never or seldom" to 5 "daily". These items covered a wide range of what German students that age typically read, including overall reading time (items a and b) and a large variety of different genres and formats (items c to l) from lowbrow and highbrow origin. We computed the arithmetic mean over all items to arrive at a single measurement (range from 1 to 5). The reliability of the generated exposure variable was fine: .74 (grade 5), .76 (grade 7) and .72 (grade 9). The distribution of this variable is visualized below in Figure 1. We would like to underscore that this measurement captures factual outside of school reading amount (and not motivation or intention). For more information on the development of the instrument (cultural capital measurement in the NEPS) refer to Goßmann (2018). The instrument has also been tested as an overall evaluation of students reading volume and approved (Locher & Pfost, 2019).

Control Variables

To approximate causal effects, we utilized a wide range of relevant control variables to account for potential confounding. According to Pearl (2009), it is relevant to include all variables that simultaneously influence both outcome variables (reading exposure and competence). Social origin is the most influential factor that can be a confounder: Socially benefited families can invest more in their children through better care, nutrition, teaching, and more, resulting in better reading competence and overall higher cognitive abilities. These families have more financial means for books and other reading materials and can provide better environments that facilitate reading, such as a room for the child to read and study in peace. Furthermore, it has been demonstrated that academic achievement and parental aspirations, which are strongly related to social origin, co-develop dynamically (Bittmann, 2022). Hence, the social origin of a family greatly influences exposure and competence, which we accounted for by including three central dimensions of social origin: the highest parental education of the family with three levels (low or intermediate education / higher education eligibility (*Abitur*) / any tertiary education). The family's financial means were measured by the total (post-tax) household income, which was logarithmized to ease statistical inference. Lastly, the parents' social status was measured by the highest reported ISEI (International Socio-Economic Index of Occupational Status), which depends on the parents' occupation in the labor market (Ganzeboom, 2010). The ISEI ranged from 11 (cleaner) to 88 (judge), and higher values indicated a higher social status. By including these measures, which capture various dimensions of social origin, we are confident that social origin is well-measured.

We added further control variables, such as gender and age of the student, which is computed for January 1, 2011 (grade 5). Older students usually have a higher reading ability as they are more mature. The number of siblings is relevant as in families with more children, parents need to divide their attention and resources, which affects how much is invested in a specific child of the family. The type of school track attended in grade 5 (academic track (*Gymnasium*) vs. any other school track) is highly relevant, since track choice is related to both primary (differences in

FIGURE 1
Distribution of Key Constructs (Competence and Exposure) by Gender and School Grade



Note. NEPS SC3, imputed data (N = 5193; M = 30)

academic performance) but also secondary effects (differences in social origin). Hence, average differences of performance are large between the academic and non-academic tracks and learning environments can differ drastically (Traini et al., 2021). We also controlled for whether a student has been diagnosed with dyslexia as this affects both reading ability but also the interest in reading; whether the student lived in a single-parent family or not as this is also related to family resources and whether one or two parents take care of a child; whether the student lived in West or East Germany due to different structural and institutional differences that remain even decades after the German reunification; the migration background of the family (yes if at least one parent was born abroad, no otherwise). In addition, parental aspirations were measured in the student questionnaire ("Which highest school-leaving qualification would your parents wish for you?"), which is either high (*Abitur*) or low (any other qualification). Recent research has outlined that measuring perceived aspirations is highly relevant as they can work differently from factual parental aspirations (Schörner & Bittmann, 2023). Lastly, we included the number of books in the household ("Around how many books do you have at home? Do not count magazines, newspapers or your textbooks."). The response scale ranged from 1 (0 to 10 books) to 6 (more than 500 books). This variable indicated how interested the family is, on average, in reading and whether the child as potential access to many books and reading material or rather not. Summarized, we argue that including these control variables makes our results more robust and minimizes the biasing influence of confounders to approximate causal effects as much as possible. If a control variable was potentially time-varying, such as the income of the family, we utilized the mode or the median over all available waves, depending on the scaling.

Statistical Analyses

To answer posed research questions, we started with a descriptive summary of all key- and control variables to understand the data. We then analyzed with t-tests whether boys and girls have different reading competence and exposure. Finally, we wanted to test whether reading competence and exposure influenced each other dynamically over time and whether feedback loops were present. Classically, cross-lagged panel models (CLPM) were the first choice to model such hypothesized relationships as they account for autoregressive effects and allow crossed influences when panel data are available. However, in recent years, new research has shown that these models can have drawbacks and are sometimes unable to recover causal effects (Hamaker et al., 2015; Usami, 2021). Based on these arguments and our specific research questions, the random intercept cross-lagged panel model (RI-CLPM) is more adequate. The RI-CLPM takes out student-constant trait-like properties of the key variables and only models deviations from a student-constant mean value (of either competence or exposure). This model tests specifically: Do students with higher deviations from their long-time average reading exposure at time point t_0 have higher deviations from their long-term average reading competence level at time point t_1? In other words, do students increase their reading competence by reading more in the previous survey wave? Note how this RI-CLPM facilitates a causal interpretation of *within-person* changes, further strengthened by including relevant control variables as outlined above (Mulder & Hamaker, 2021). While by design, the RI-CLPM accounts for time-constant confounders, it assumes that the effect a potential confounder has on the outcome variables is stable over time and does not change with the survey waves (Mund et al., 2021). However, if this assumption is violated, bias can be introduced. This is the main reason for including the control variables as described above, even if they are time-constant by nature. In general, we also would like to acknowledge that the RI-CLPM is not a panacea, and there is valid criticism of the model as well (Lüdtke & Robitzsch, 2021). They even outline that the RI-CLPM is not always able to account for time-constant confounding variables, which is one of the reasons why we opted for the inclusion of relevant control variables. To account for these potential issues with the RI-CLPM, we also reported the classical CLPM with a lag-2 structure, as suggested by other researchers (Gnambs & Lockl, 2023). These models answer slightly different questions, including both within-person and between-person effects. While the RI-CLPM takes out the between-person variance and only answers how deviations from a person-centered mean develop over time, the CLPM estimates total effects. Consequently, the path coefficients in a CLPM are usually larger than the same coefficients in a RI-CLPM. CLPMs describe stability and change in individual differences, and statistically significant effects represent systematic effects on change over time (Selig & Little, 2012).

Data were prepared and analyzed using Stata 16.1 (Bittmann, 2019); the only exceptions were the (RI-)CLPMs, which were estimated with lavaan 0.6.17 in R (Rosseel, 2012). All reported RI-CLPMs were estimated using a maximum likelihood estimator. The goodness of fit of the models was checked using five indicators: chi-square test, comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Good model fit is demonstrated by a small, preferably statistically nonsignificant chi-square[3], CFI and TLI greater than .95, RMSEA less than .06, and SRMR less than .08 (Hu & Bentler, 1999). Missing data was accounted for using full information maximum likelihood (FIML) for the (RI-)CLPMs or multiple imputation with chained equations (MICE; all other analyses; 30 imputed datasets generated).

Results

Descriptive Statistics

In Table 1, we provide descriptive statistics for all variables. The average age in the sample was 11.1 years at the start of the study, and the gender composition was highly balanced. Table 2 shows a correlation matrix for the key variables (reading competence and exposure) over all three points in time. Interestingly, reading exposure and competence were not correlated in grade 5, which is at the very start of secondary education. However, this association became much stronger in the subsequent waves. Next, we show graphically how the key variables are distributed by gender (Figure 1).

Reading competence fits a normal distribution very well. Average competence increases over time as the curves from grade 7 and grade 9 are shifted to the right. This makes sense as the tests are constructed in such a way as to account for increasing reading competence over time. Regarding reading exposure, the graphs are slightly skewed. Interestingly, reading exposure slightly decreases over time, which holds for both genders. This means that the students in the sample tend to read less often as they age.

Testing for Gender Differences

Before discussing the main models of the study, we would like to demonstrate how reading competence and exposure differ by gender. This has two purposes. First, this tests whether the internationally well-established gender differences in reading ability can be reproduced with the NEPS data. Second, if gender differences are revealed,

TABLE 1
Descriptive Statistics

	Mean/Share	SD	Median	Min	Max	Share missing
Competence grade 5	.00	1.00	.00	−3.70	3.14	0
Competence grade 7	.55	1.08	.55	−2.68	4.58	.26
Competence grade 9	1.01	.88	.99	−1.65	4.95	.45
Exposure grade 5	2.43	.74	2.33	1	5	.03
Exposure grade 7	2.14	.68	2.08	1	5	.19
Exposure grade 9	1.97	.59	1.92	1	5	.29
Female student	.48			0	1	0
Living in East Germany	.15			0	1	0
Single parent	.24			0	1	.26
Total household income (log)	8.07	.48	8.09	5.98	10.0	.35
Age of student in January 2011	11.1	.50	11.0	10.0	13.8	<.01
Migration background	.21			0	1	.26
Parental education level						.26
Low/intermediate	.48			0	1	
Higher education eligibility	.22			0	1	
Tertiary	.30			0	1	
Number of siblings	1.11	.95	1	0	10	.04
Dyslexia diagnosis	.073			0	1	.26
Number of books in HH	4.21	1.44	4	1	6	<.01
Highest parental ISEI	53.8	20.4	54.5	11.7	89.0	.27
Academic track (grade 5)	.44			0	1	<.01
High perceived parental aspirations	.68			0	1	.08

Note. NEPS SC3, imputed data (N = 5193; M = 30). HH = household.

TABLE 2
Correlation Matrix for Reading Competence and Reading Exposure

	Competence 5	Competence 7	Competence 9	Exposure 5	Exposure 7	Exposure 9
Competence 5	–					
Competence 7	.62***	–				
Competence 9	.60***	.64***	–			
Exposure 5	.00	.01	.03*	–		
Exposure 7	.16***	.18***	.16***	.36***	–	
Exposure 9	.17***	.22***	.21***	.30***	.47***	–

Note. NEPS SC3, imputed data (N = 5193; M = 30).
*p < .05, ***p < .001.

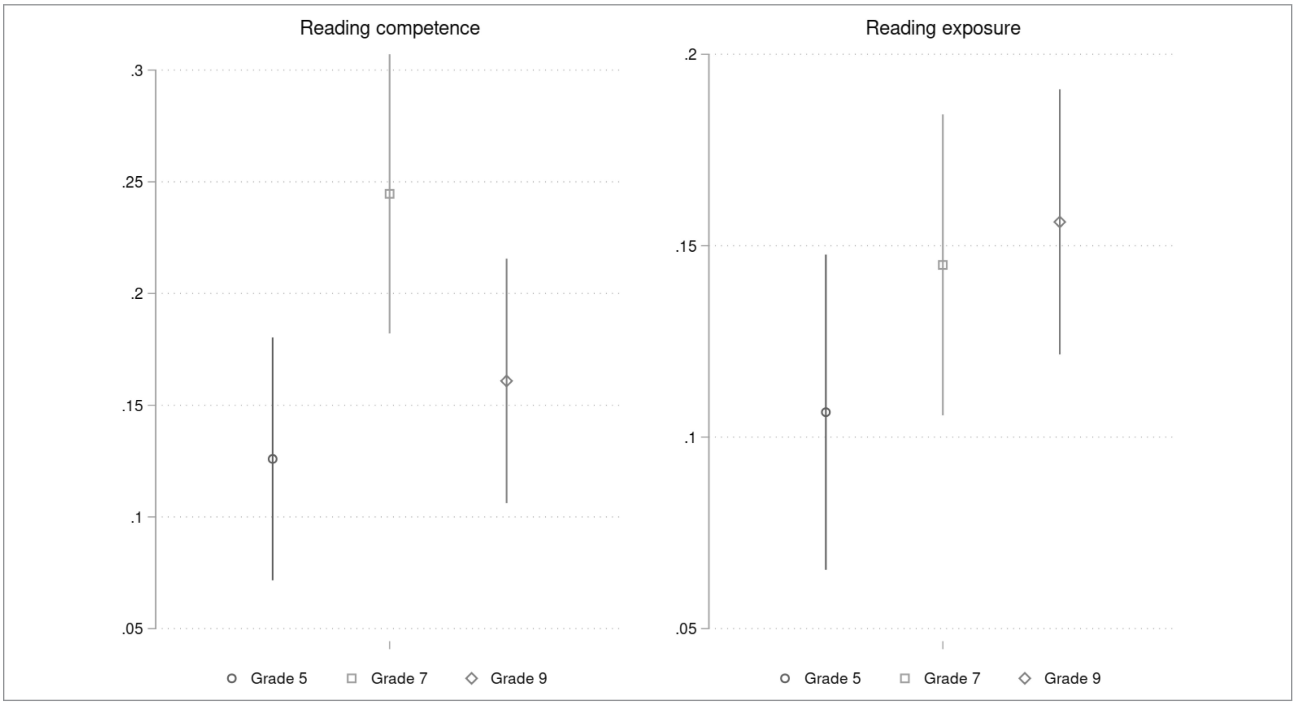
this is good argument for not only including gender as a control variable but also compute a stratified RI-CLPM to test whether the crossed effects between reading competence and exposure are different for the two genders. Statistically, we conducted t-tests for the two key constructs at all three points in time. Note that these analyses were descriptive and did not include any control variables. For convenience, the results were summarized graphically. For inference, 95% confidence intervals were included. Concerning gender differences in reading competence (Figure 2, left side), we see that girls have a statistically significant advantage over boys. This advantage tends to grow slightly in grade 7 yet goes back to the initial level in grade 9.Girls also have an advantage concerning reading exposure, meaning they tend to read more than boys (Figure 2, right side). This difference even grows slightly over time. For all coefficients, the difference between the genders is statistically significant on the 5% level since the confidence intervals never touch zero. Regarding effect sizes, the largest gender differences are about 25% of a standard deviation. To summarize, these findings align with previous studies that report the same advantages for girls for reading competence and exposure (Schleicher, 2019; Van Hek et al., 2019).

RI-CLPM

Next, we computed the main model of interest, the RI-CLPM. This model can tell us whether reading competence and exposure develop dynamically over time and influence each other. We are confident that this model can recover causal effects by including a large set of relevant control variables (as theoretically outlined above). We first evaluated the model fit to see whether the specified model fits the data well. $\chi^2(99)=7819$, $p<.0001$; CFI = .999; TLI = .999; RMSEA = .030, 90% confidence interval [.011, .055]; SRMR = .002. The CFI and TLI are excellent and indicate a very good model fit. The RMSEA is also below .05, which indicates a fine model fit; the SRMR is also excellent. While a statistically nonsignificant chi-squared value is preferred, it is well known that given a large sample size, such as in our case ($N=5193$), a small p-value is very likely and should not be used to judge the overall model fit (Schermelleh-Engel et al., 2003; Vandenberg, 2006). Summarized, these results indicate that the model fit is fine.

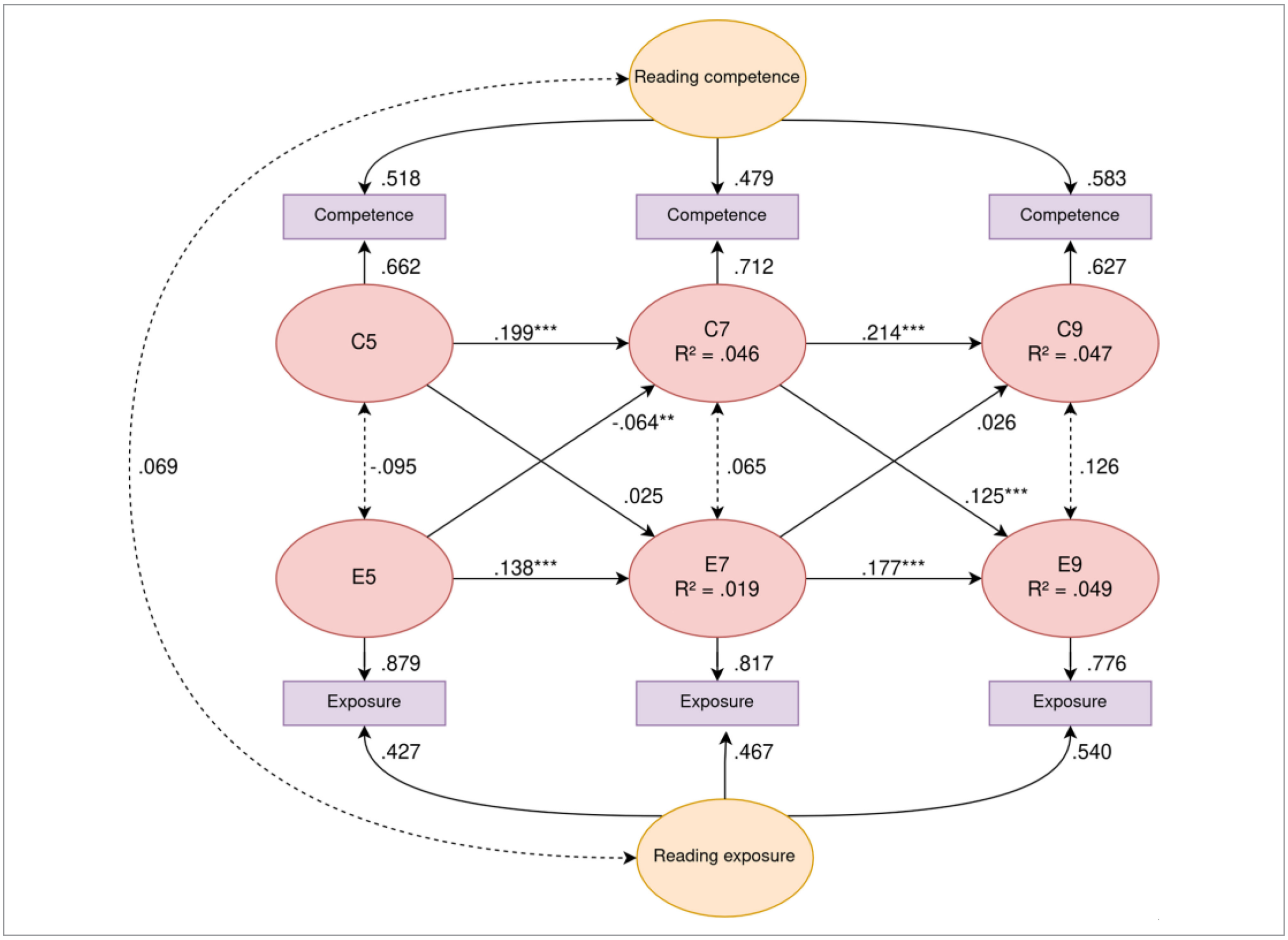
We continue with the path coefficients. These can tell us whether crossed effects and feedback processes are present. For a convenient interpretation, we summarized the results for the RI-CLPM in Figure 3. We see that the autocorrelations are always statistically highly significant. As the competences measures were z-standardized, we can interpret the values as follows: if the competence of a student increases by one standard deviation in grade 5, the competence in grade 7 increases by about .2 standard deviations. This makes sense as it outlines that initial competence (exposure) predicts future competence (exposure). We did not find any evidence regarding exposure's positive influence on competence. The effect of exposure$_5$ on competence$_7$ is even slightly negative and close to zero, which hints for a null-finding. The effect of exposure$_7$ on competence$_9$ is not statistically significant at all. Overall, the main conclusion is that in the RI-CLPM, exposure has no positive effect on competence. Regarding the other

FIGURE 2
Gender Difference of Competence and Exposure by School Grade



Note. NEPS SC3, imputed data ($N=5193$; M = 30). A positive number indicates an advantage for girls. 95% confidence intervals included.

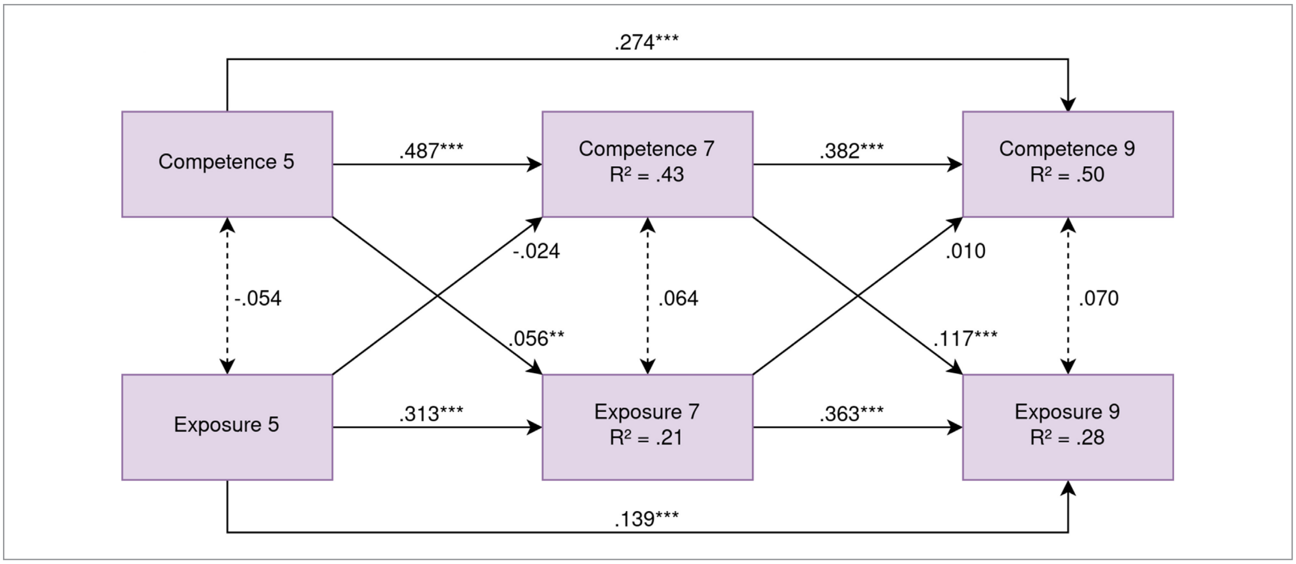
FIGURE 3
Random Intercept Cross-Lagged Panel Model for the Development of Reading Competence (C_t) and Exposure (E_t)



Note. NEPS SC3 (N=5193). Control variables: gender, age, place of residence, parental income, parental education, parental ISEI, single parent, migration background, number of siblings, dyslexia diagnosis, parental aspirations, school track, and number of books in the household. Standard errors clustered by school. Dashed lines represent covariances. Reported are standardized coefficients (ß). The between-subject part (the top and bottom ovals) captures students' general level. The within-subject part (ovals in the middle) shows the within-student associations. R^2 values indicate only explained within-person variance. *p < .05, **p < .01, ***p < .001.

crossed effects from competence on exposure, there is only evidence from competence$_7$ on exposure$_9$. This effect is statistically highly significant. The interpretation is as follows: if the standard deviation of a student increases by one standard deviation in grade seven, the reading exposure increases by about .125 standard deviations in grade nine. Note that the association between the two random intercepts is weak (.069), which means that most of the trait associations are at the within-student level. When interpreting reported R^2 values as a measurement of explained variance, keep in mind that the RI-CLPM only accounts for within-person changes and all between-person differences are partialled out, not contributing to the computation of the explained variance. The values are further decreased by the usage of control variables since the effects that are due to these variables are also not accounted for by the R^2 values in a RI-CLPM. This is different in the following CLPM, where all variables and between-person effects contribute to R^2; the values are much larger in this model.As a robustness check, we computed a standard CLPM with a lag-2 structure (meaning that variables at grade 5 were allowed to influence variables at grade 9 directly). Others have pointed out that such a model is an alternative to test causal influences, especially when relevant control variables are available. The model fit of this CLPM was excellent: $\chi^2(99)=7819$, $p<.0001$; CFI = 1.00; TLI = 1.00; RMSEA = .000, 90% confidence interval [.000, .018]; SRMR = .001. However, we want to emphasize that model fit is not a sensible way to decide whether the RI-CLPM or the CLPM provides "better" results since the interpretation of the path coefficients differs due to the model parameterization. We show the main results graphically in Figure 4. Again, all autocorrelations are statistically highly significant, which also holds for the lag-2 effects (crossed lag-2 effects are not included since they are usually not sensible to interpret). Regarding

FIGURE 4
Cross-Lagged Panel Model for the Development of Reading Competence and Exposure



Note. NEPS SC3 (N = 5193). Control variables: gender, age, place of residence, parental income, parental education, parental ISEI, single parent, migration background, number of siblings, dyslexia diagnosis, parental aspirations, school track, and number of books in the household. Standard errors clustered by school. Dashed lines represent covariances. Reported are standardized coefficients (β). R^2 values indicate explained within- and between-person variance. *$p < .05$, **$p < .01$, ***$p < .001$.

the crossed effects, we find no evidence that exposure influences competence. However, there are statistically significant effects for competence on exposure for both measurement points. This is in line with the findings of the RI-CLPM (the main difference is that the effect of competence$_5$ on exposure$_7$ is statistically significant in the CLPM only).To summarize the findings so far, we conclude that reading exposure has no positive influence on reading competence. While the RI-CLPM even shows a small negative effect of exposure$_5$ on competence$_7$ (which is potentially a null-finding since it is very close to zero), we do not believe that this coefficient is robust since it is not statistically significant at all in the CLPM. Regarding the effects of competence on exposure, we do find some evidence in both models. Therefore, reading competence has at least some causal influence on reading exposure for German secondary school students.

Gender Differences

Finally, we would like to test whether the same conclusions hold for boys and girls or whether they have different relationships between reading competence and exposure. To answer this question, we computed a multigroup RI-CLPM and statistically test for differences in the path coefficients. First, we provide information on the general model fit: $\chi^2(186) = 7794$, $p < .0001$; CFI = .999; TLI = .999; RMSEA = .033, 90% confidence interval [.010, .059]; SRMR = .002. These results indicate a fine model fit. For simplicity, we provide the standardized path coefficients for boys and girls in Table 3 and their differences (the same control variables are applied as for the RI-CLPM described above).

As the results indicate, the results for boys and girls are highly similar. Not a single difference is statistically significant on the 5% level. We can conclude from these findings that the general dynamic relationship between reading competence and exposure is not gender dependent, even if gender differences in competence and exposure were

TABLE 3
Standardized Path Coefficients for the Multigroup RI-CLPM

Pathway	Coef. Boys	Coef. Girls	Difference	SE of difference
$C_5 \to C_7$	.183	.212	−.028	.056
$C_7 \to C_9$	.236	.171	.064	.061
$E_5 \to E_7$	.120	.159	−.038	.056
$E_7 \to E_9$	.178	.171	.007	.071
$C_5 \to E_7$	.019	.041	−.022	.055
$E_5 \to C_7$	−.052	−.077	.025	.042
$C_7 \to E_9$	.128	.133	−.005	.056
$E_7 \to C_9$	−.026	.094	−.120	.067
N	2689	2504		

Note. NEPS SC3 (*N* = 5193). No difference is statistically significant on the 5% level. “Difference” can slightly deviate from the difference between Boys and Girls due to rounding. C_t represents reading competence at the given grade, E_t represents reading exposure.

found. As the number of both male and female students in the sample is large, we argue that these results are probably rather stable.

Discussion

Our findings outline that there are probably no crossed effects between reading exposure and reading. While the correlation matrix demonstrates that reading competence and exposure are correlated, the advanced models report that this association is probably only partial causal. One link we can establish rather confidently is the positive effect of reading competence on exposure. The CLPM is clear on this aspect since for both points in time, from grade 5 to 7 and from grade 7 to 9, this model reports statistically significant effects, which even become larger over time. Note that this model includes both within- and between-student effects. When we focus on the RI-CLPM, which is closer to a causal interpretation, we come to the same conclusions only for older students (from grades 7 to 9). At this point, it is obvious that students who increase their reading competence also increase their reading exposure in the following survey wave. For the earlier measurement, this effect is tentative and statistically insignificant. Overall, our findings align with previous studies, which conclude that better competence leads to higher exposure (Erbeli et al., 2020). While these finding also makes sense from a theoretical point of view as students who can decode words more easily and quickly enjoy reading more and potentially have a higher motivation to read, we note that our effect sizes are much smaller than the results from a comparable study, which also utilized the RI-CLPM (Van Bergen et al., 2021). We assume that these differences can be explained by the usage of a large set of relevant control variables in our study. And indeed, when we estimated our RI-CLPM without any controls, the effects became stronger (not reported). As most of our control variables are time-constant by nature, this should not happen if the effect of controls on the outcome variables is identical for all survey waves. Apparently, this is not the case and the inclusion of control variables is necessary to approximate causal effects. We would also like to underscore that the inclusion of controls is unlikely to result in overcontrol bias since our controls are antecedents of both reading achievement and exposure, and neither consequences nor mediators. For example, parents neither change their migration background nor their level of education if their child increases his or her reading ability. Therefore, we are confident that the introduction of relevant control variables, even in the RI-CLPM, can be necessary and beneficial. It might be possible that previous studies report over-estimated crossed effects due to the omission of control variables. Theoretically, there are many arguments available why better readers tend to read more. Referring to the SVR-model, one can assume that better readers require less cognitive capacity to decode and comprehend texts, which makes the entire process for them more enjoyable, resulting in more time spent reading. Our analyses support the view that this is indeed a causal relationship, since the number of books at home was controlled for. This means that results are independent of whether children grew up in an environment that offered little or many books. One conclusion of this finding is that open access to reading material for skilled readers probably increases the time a child will spend reading.

When the focus is on the reciprocal pathway, from reading exposure to reading competence, we do not find any evidence. The statistical models agree and show no positive influence of reading exposure on competence. The RI-CLPM even reports a very small yet statistically significant *negative* effect for the earlier measurement point. However, since this effect vanishes for the second point in time and the CLPM does not support this conclusion, it should not be overstated as it is trivially small. Our findings align with previous studies that only find effects in one direction and not from exposure to competence (Altamura et al., 2023; Harlaar et al., 2011). Our findings are valid for German secondary school students and cover the age range from 11 to 15 years. This finding indicates that simply providing plenty of reading material to kids will not automatically result in better reading performance. Access alone is potentially a requirement, yet not sufficient to boost reading performance, and other aspects, such as continuously providing support and guidance, potentially through parents and teachers, is necessary for advancing reading competence in children. Especially parents should be aware of the fact that helping their children read and reading together are advised and more than access to books and reading material is required to facilitate a positive development of reading performance.

Regarding gender differences in the process, we do not find any effects at all. Our results outline that the coefficients are usually highly similar for boys and girls, and no statistically significant difference is detectable. The only potential exception is the effect of reading exposure in grade 7 on reading competence in grade 9. Here, boys have a small negative effect, but girls have a large positive effect. This could hint at a potential effect of reading more on reading competence that is only present for girls but not boys. However, given that our sample size is already large, this gender difference is very small and hard to detect. Consequently, we suggest not to overstate it, especially since it is not visible at all earlier, where the effects of boys and girls are highly similar. The conclusion is, in general, that the dynamic interplay between reading competence and reading exposure is the same for boys and girls, even if girls are better and avid readers, as is not only visible in our data but in the literature in general. Teachers and parents should be aware of this fact and not treat boys and girls differently, even if girls are better readers, on average.

Study Characteristics and Limitations

While our study provides compelling evidence, it is also necessary to acknowledge its limitations. First, while we can provide longitudinal data, the key constructs were measured with a one-year gap. What happens to reading competence and exposure in school grades 6 and 8 is not accessible to us. As reading is a process that develops continuously over time, it would be desirable to have more measurement points to gain more insight. However, this gap should not necessarily introduce bias even if diminishing insight. Second, reading exposure was measured using a questionnaire, which can introduce measurement error since actual reading behavior was not directly observed. However, this is the only measurement option feasible for a pen-and-paper survey, and this limitation must be accepted. While it can also introduce a social desirability bias, and students might want to overestimate their reading exposure, since the data were longitudinal, this should not heavily affect the outcomes since *changes* from the student-centered mean are relevant and not absolute levels. Third, we modeled all constructs as manifest and not latent ones, which is also an option in structural equation modeling. However, since the scaling and linking procedures of the reading competence tests were complex and carried out by the NEPS research department, which provided WLE (weighted likelihood estimates) scores to researchers, we decided to model all constructs as manifest ones. This handling is similar to many previous studies and should not introduce bias.

Fourth, the number of usable cases declined over time due to panel attrition. This concerns especially the competence tests, as the share of missingness increases from zero to about 45% over five school years. This is a common problem in longitudinal surveys and is difficult to combat. While this issue decreases power due to the lower information, it should not introduce selection or survivorship bias. As missing data was imputed (using either FIML or MICE, which are asymptotically equivalent), selective dropout was accounted for if the data were missing at random (MAR). Dropout could be explained largely by social origin, and students from the lower social strata have a higher propensity to drop out. This aspect was covered by including the various measurements of social origin in our models. In addition, competencies were rather stable over time, and earlier measurements were highly predictive of later ones. As every student had a valid measurement in the first survey wave, this information was greatly relevant for predicting potential missing values. Overall, panel attrition is not an overwhelmingly problematic issue, and our sample is still very large compared to many previous studies with only a few hundred cases. Fifth, while the NEPS data included students from all over Germany, it is difficult to prove that the sample is representative of the overall German student population as no official statistics were surveyed, which could serve as a benchmark. Nevertheless, we hope that by providing descriptive statistics for many variables, readers are empowered to judge the external validity of our study. Sixth and final, the reading exposure measurement mostly focused on print media such as books and magazines and less on digital reading material, which is becoming more and more relevant for students and society. However, reading printed material is still a major competence, and books and other material reading products will not vanish anytime soon. Therefore, our findings still have major practical implications.

Conclusion

Our results indicate that increasing reading competence leads to increased reading exposure, at least for German secondary school students. This finding and its causal interpretation are especially indicated by the RI-CLPM, which shows that the students who positively deviate from their long-term reading competence also increase their reading exposure in the subsequent measurement wave. This finding holds for both boys and girls, despite the fact that girls are, on average, better and more avid readers. These findings support multiple conclusions. First, as there is no indication that more reading exposure leads to better reading skills, this hints for the necessity of continuous instruction in reading skill development. Even after students have learned the basics of reading in primary school, simply immersing students in print-rich environments or giving them more time to read does not automatically lead to better reading competence. While providing these environments and giving them enough time for reading is clearly not harmful, additional reading instruction is probably required to hone reading skills. Especially parents and teachers should be aware of this and help their children with reading and support them whenever necessary in addition to providing reading material alone. We assume that the effects this study recovers are smaller than previously reported ones due to a focus on the approximation of causal effects and the inclusion of control variables. Remember that many previous studies have also not found such causal effects, and the evidence is mixed. However, the current study only includes students aged 11 to 15, and the effects might differ for younger or older students. Still, our results provide compelling evidence based on large-scale German panel data.

A second main conclusion of our study is that the dynamic feedback process between reading competence and exposure is not stratified by gender, meaning that the same mechanisms and results apply to both boys and girls. These results hold, even though we can demonstrate that girls are, on average, more avid and better readers, which is also the obvious conclusion of many previous studies. Based on these findings, no special or different treatment of the genders is necessary to facilitate the process of gaining reading competence. The only exception is that older girls potentially profit more from a higher reading exposure than boys; however, this finding is rather unstable and should not be overstated. Summarized, our study has provided new evidence on the dynamic co-development of reading competence and exposure for secondary school students. Extending this research focus to older and younger students would be desirable, as our current results cannot be generalized to these other groups.

Acknowledgments

Special thanks to Elif Sari and Karin Gehrer for answering various questions about measuring reading exposure and reading competence in the NEPS. I also want to thank the two anonymous reviewers and the journal editor for their very helpful comments and suggestions. Open Access funding enabled and organized by Projekt DEAL.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no conflict of interest.

Data Availability Statement

Data are available after registration as a researcher from https://doi.org/10.5157/NEPS:SC3:12.1.0.

ENDNOTES

[1] This paper uses data from the National Educational Panel Study (NEPS; see Blossfeld and Roßbach (2019)). The NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi, Germany) in cooperation with a nationwide network.

[2] Note that an additional reading test has been conducted in survey wave 12. However, extending the analyses to more than three waves is infeasible since reading exposure is no longer surveyed in these later waves.

[3] We would like to emphasize that the chi-square value can be rather meaningless when the sample size is large. In large-scale samples, a statistically significant result is almost always produced. Hence, the other statistics might be more relevant to judge the model fit.

REFERENCES

Aarnoutse, C., & Van Leeuwe, J. (1998). Relation between Reading comprehension, vocabulary, Reading pleasure, and Reading frequency. *Educational Research and Evaluation*, *4*(2), 143–166. https://doi.org/10.1076/edre.4.2.143.6960

Allington, R. L., & McGill-Franzen, A. M. (2021). Reading volume and Reading achievement: A review of recent research. *Reading Research Quarterly*, *56*(S1), S231–S238. https://doi.org/10.1002/rrq.404

Altamura, L., Vargas, C., & Salmerón, L. (2023). Do new forms of Reading pay off? A meta-analysis on the relationship between leisure digital Reading habits and text comprehension. *Review of Educational Research*. https://doi.org/10.3102/00346543231216463

Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in Reading and how children spend their time outside of school. *Reading Research Quarterly*, *23*(3), 285–303. https://doi.org/10.1598/RRQ.23.3.2

Becker, M., McElvany, N., & Kortenbruck, M. (2010). Intrinsic and extrinsic reading motivation as predictors of reading literacy: A longitudinal study. *Journal of Educational Psychology*, *102*(4), 773–785. https://doi.org/10.1037/a0020084

Bittmann, F. (2019). *Stata: A really short introduction* (1st ed.). De Gruyter Oldenbourg.

Bittmann, F. (2022). Investigating the co-development of academic competencies and educational aspirations in German primary education. *Frontiers in Education*, *7*, 923361. https://doi.org/10.3389/feduc.2022.923361

Blossfeld, H., & Roßbach, H. (2019). *Education as a lifelong process: The German National Educational Panel Study (NEPS)* (Vol. 3, 2nd edn, Edition ZfE ed.). VS Verlag für Sozialwissenschaften.

Brüderl, J., & Ludwig, V. (2015). Fixed-effects panel regression. *The sage handbook of regression analysis and causal inference* (pp. 327–357). Sage.

Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. *Journal of Educational Psychology*, *82*(4), 733–740. https://doi.org/10.1037/0022-0663.82.4.733

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934–945. https://doi.org/10.1037/0012-1649.33.6.934

Ditton, H. (Ed.). (2007). *Kompetenzaufbau und Laufbahnen im Schulsystem: Ergebnisse einer Längsschnittuntersuchung an Grundschulen*. (1. Aufl). Waxmann.

Erbeli, F., Van Bergen, E., & Hart, S. A. (2020). Unraveling the relation between Reading comprehension and print exposure. *Child Development*, *91*(5), 1548–1562. https://doi.org/10.1111/cdev.13339

Fischer, L., Gnambs, T., Rohm, T., & Carstensen, C. H. (2019). Longitudinal linking of Rasch-model-scaled competence tests in large-scale assessments: A comparison and evaluation of different linking methods and anchoring designs based on two tests on mathematical competence administered in grades 5 and 7. *Psychological Test and Assessment Modeling*, *61*(1), 37–64.

Fischer, L., Rohm, T., Gnambs, T., & Carstensen, C. H. (2016). *Linking the data of the competence tests*. NEPS Survey Paper, 1. https://www.lifbi.de/Portals/2/Survey%20Papers/SP_I.pdf

Ganzeboom, H. B. (2010). A new international socio-economic index (ISEI) of occupational status for the international standard classification of occupation 2008 (ISCO-08) constructed with data from the ISSP 2002-2007. Annual conference of international social survey Programme. Lisbon, 1 http://www.harryganzeboom.nl/pdf/2010%20-%20Ganzeboom%20-%20A%20New%20International%20Socio-Economic%20Index%20ISEI%20of%20occupational%20status%20for%20the%20International%20Standard%20Classification%20of%20Occupation.pdf

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.

Gnambs, T., & Lockl, K. (2023). Bidirectional effects between reading and mathematics development across secondary school. *Zeitschrift für Erziehungswissenschaft*, *26*(2), 345–371. https://doi.org/10.1007/s11618-022-01108-w

Goßmann, F. (2018). *Measuring Cultural Capital in the NEPS* (*NEPS Survey Paper No. 48*). Leibniz Institute for Educational Trajectories, National Educational Panel Study. https://www.neps-data.de/Portals/0/Survey%20Papers/SP_XLVIII.pdf

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Harlaar, N., Deater-Deckard, K., Thompson, L. A., DeThorne, L. S., & Petrill, S. A. (2011). Associations between Reading achievement and independent Reading in early elementary school: A genetically informative cross-lagged study: Reading achievement and independent Reading. *Child Development*, *82*(6), 2123–2137. https://doi.org/10.1111/j.1467-8624.2011.01658.x

Hernán, M. A. (2018). The C-word: Scientific euphemisms do not improve causal inference from observational data. *American Journal of Public Health*, *108*(5), 616–619.

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

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jabbar, A., & Warraich, N. F. (2023). Gender differences in leisure reading habits: A systematic review of literature. *Global Knowledge, Memory and Communication*, *72*(6/7), 572–592. https://doi.org/10.1108/GKMC-12-2020-0200

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*(4), 437–447.

Kunter, M., Schümer, G., & Artelt, C. (2002). *PISA* 2000*: Dokumentation der Erhebungsinstrumente*. 410.

Lüdtke, O., & Robitzsch, A. (2021). *A Critique of the Random Intercept Cross-Lagged Panel Model* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/6f85c

Leppänen, U., Aunola, K., & Nurmi, J. (2005). Beginning readers' reading performance and reading habits. *Journal of Research in Reading*, *28*(4), 383–399. https://doi.org/10.1111/j.1467-9817.2005.00281.x

Lietz, P. (2006). A meta-analysis of gender differences in Reading achievement at the secondary school level. *Studies in Educational Evaluation*, *32*(4), 317–344. https://doi.org/10.1016/j.stueduc.2006.10.002

Locher, F. M., & Pfost, M. (2019). Measuring reading volume in large-scale assessments: A comparison of an overall evaluation and a differentiated evaluation relating different text types. *Diagnostica*, *65*(1), 26–36.

Logan, S., & Johnston, R. (2010). Investigating gender differences in reading. *Educational Review*, *62*(2), 175–187. https://doi.org/10.1080/00131911003637006

Lucas, R. E. (2023). Why the cross-lagged panel model is almost never the right choice. *Advances in Methods and Practices in Psychological Science*, *6*(1), 25152459231158378. https://doi.org/10.1177/25152459231158378

Marszalek, J. M., Barber, C., Kohlhart, J., & Cooper, B. H. (2011). Sample size in psychological research over the past 30 years. *Perceptual and Motor Skills*, *112*(2), 331–348. https://doi.org/10.2466/03.11.PMS.112.2.331-348

McGue, M., Osler, M., & Christensen, K. (2010). Causal inference and observational research: The utility of twins. *Perspectives on Psychological Science*, *5*(5), 546–556. https://doi.org/10.1177/1745691610383511

Millard, E. (1998). Tackling gender differences in reading. Gender in the Secondary Curriculum: Balancing the Books. 43.

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

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). Cambridge University Press.

Morley, R. (2005). Can we generalise from findings in twins? *Paediatric and Perinatal Epidemiology*, *19*(s1), 54–59. https://doi.org/10.1111/j.1365-3016.2005.00610.x

Mulder, J. D., & Hamaker, E. L. (2021). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(4), 638–648. https://doi.org/10.1080/10705511.2020.1784738

Mund, M., Johnson, M. D., & Nestler, S. (2021). Changes in size and interpretation of parameter estimates in within-person models in the presence of time-invariant and time-varying covariates. *Frontiers in Psychology*, *12*, 666928. https://doi.org/10.3389/fpsyg.2021.666928

NEPS Network. (2024). *National educational panel study, scientific use file of starting cohort grade 5*. Leibniz Institute for Educational Trajectories (LIfBi). https://doi.org/10.5157/NEPS:SC3:13.0.0

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, *3*, 96–146. https://doi.org/10.1214/09-SS057

Pohl, S. (2013). Longitudinal Multistage Testing. *Journal of Educational Measurement*, *50*(4), 447–468. https://doi.org/10.1111/jedm.12028

Pohl, S., & Carstensen, C. H. (2013). *Scaling of competence tests in the National Educational Panel Study—Many questions, some answers, and further challenges*. https://doi.org/10.25656/01:8430

Rohrer, J. M., & Murayama, K. (2023). These are not the effects you are looking for: Causality and the within-/between-persons distinction in longitudinal data analysis. *Advances in Methods and Practices in Psychological Science*, *6*(1). https://doi.org/10.1177/25152459221140842

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Schleicher, A. (2019). *PISA 2018: Insights and interpretations* (p. 64). Oecd Publishing.

Schörner, K., & Bittmann, F. (2023). Children's aspirations, their perceptions of parental aspirations, and parents' factual aspirations—gaining insights into a complex world of interdependencies. *European Sociological Review*. https://doi.org/10.1093/esr/jcad074

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In *Handbook of developmental research methods* (pp. 265–278). The Guilford Press.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*(2), 95–129. https://doi.org/10.1006/jecp.1998.2481

Stanovich, K. E. (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education*, *189*(1–2), 23–55.

Thums, K., Artelt, C., & Wolter, I. (2021). Reading for entertainment or information reception? Gender differences in reading preferences and their impact on text-type-specific reading competences in adult readers. *European Journal of Psychology of Education*, *36*(2), 339–357. https://doi.org/10.1007/s10212-020-00486-1

Tipton, E., Hallberg, K., Hedges, L. V., & Chan, W. (2017). Implications of small samples for generalization: Adjustments and rules of thumb. *Evaluation Review*, *41*(5), 472–505. https://doi.org/10.1177/0193841X16655665

Torppa, M., Niemi, P., Vasalampi, K., Lerkkanen, M., Tolvanen, A., & Poikkeus, A. (2020). Leisure Reading (but not any kind) and Reading comprehension support each other—A longitudinal study across grades 1 and 9. *Child Development*, *91*(3), 876–900. https://doi.org/10.1111/cdev.13241

Traini, C., Kleinert, C., & Bittmann, F. (2021). How does exposure to a different school track influence learning progress? Explaining scissor effects by track in Germany. *Research in Social Stratification and Mobility*, *100625*, 100625. https://doi.org/10.1016/j.rssm.2021.100625

Usami, S. (2021). On the differences between general cross-lagged panel model and random-intercept cross-lagged panel model: Interpretation of cross-lagged parameters and model choice. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(3), 331–344. https://doi.org/10.1080/10705511.2020.1821690

Van Bergen, E., Snowling, M. J., De Zeeuw, E. L., Van Beijsterveldt, C. E. M., Dolan, C. V., & Boomsma, D. I. (2018). Why do children read more? The influence of reading ability on voluntary reading practices. *Journal of Child Psychology and Psychiatry*, *59*(11), 1205–1214. https://doi.org/10.1111/jcpp.12910

Van Bergen, E., Vasalampi, K., & Torppa, M. (2021). How are practice and performance related? Development of Reading from age 5 to 15. *Reading Research Quarterly*, *56*(3), 415–434. https://doi.org/10.1002/rrq.309

Van Hek, M., Buchmann, C., & Kraaykamp, G. (2019). Educational systems and gender differences in Reading: A comparative multilevel analysis. *European Sociological Review*, *35*(2), 169–186. https://doi.org/10.1093/esr/jcy054

Vandenberg, R. J. (2006). Introduction: Statistical and methodological myths and urban legends: Where, pray tell, did they get this idea? In *Organizational research methods* (Vol. 9, pp. 194–201). Sage Publications Sage CA.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., Carstensen, C. H., & Lockl, K. (2019). *Development of competencies across the life course* (pp. 57–81). Education as a Lifelong Process.

Submitted March 29, 2024
Final revision received May 31, 2024
Accepted July 1, 2024

Felix Bittmann is a postdoctoral research fellow at the Leibniz Institute for Educational Trajectories (LifBi) in Bamberg, Germany; email felix.bittmann@lifbi.de. He is a sociologist by training, and his research focuses mostly on social inequality in the German educational system. He is deeply interested in modern quantitative and causal methods and works on various topics from the social sciences.