



Modeling Interactions Between Multivariate Learner Characteristics and Interventions: a Person-Centered Approach

Leonard Tetzlaff¹ · Peter Edelsbrunner^{2,3} · Alexandra Schmitterer⁴ · Ulrike Hartmann¹ · Garvin Brod^{1,5}

Accepted: 3 November 2023 / Published online: 28 November 2023
© The Author(s) 2023

Abstract

Demonstrating the differential effectiveness of instructional approaches for learners is difficult because learners differ on multiple dimensions. The present study tests a person-centered approach to investigating differential effectiveness, in this case of reading instruction. In $N = 517$ German third-grade students, latent profile analysis identified four subgroups that differed across multiple characteristics consistent with the *simple view of reading*: poor decoders, poor comprehenders, poor readers, and good readers. Over a school year, different instructional foci showed differential effectiveness for students in these different profiles. An instructional focus on vocabulary primarily benefited good readers at the expense of poor decoders and poor comprehenders, while a focus on advanced reading abilities benefitted poor comprehenders at the expense of poor decoders and good readers. These findings are in contrast to those obtained by multiple regression, which, focusing on only one learner characteristic at a time, would have suggested different and potentially misleading implications for instruction. This study provides initial evidence for the advantages of a person-centered approach to examining differential effectiveness.

Keywords Differential effectiveness · Person-centered analysis · Simple view of reading · Aptitude-treatment interactions · Latent profile analysis

✉ Leonard Tetzlaff
l.tetzlaff@dipf.de

¹ DIPF | Leibniz Institute for Research and Information in Education, Rostocker Straße 6, 60323 Frankfurt, Germany

² ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland

³ Ludwigs-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 München, Germany

⁴ University of Paderborn, Warburger Straße 100, 33098 Paderborn, Germany

⁵ Goethe Universität Frankfurt, Theodor-W.-Adorno-Platz 1, 60323 Frankfurt, Germany

When teachers engage in differentiated instruction, they often begin by intuitively grouping students based on multiple characteristics that they consider relevant prerequisites for learning (Corno, 2008). For example, they might view some students as smart but lazy and others as motivated learners with weak self-regulation. Based on their experience and intuitive diagnosis, they use such categories as a heuristic for adapting their teaching to learners' individual needs.

In contrast to this common educational practice in which educators consider multiple characteristics for each student, most research investigating the differential effectiveness of educational interventions focuses on only one characteristic (e.g., intelligence or prior knowledge; Kalyuga, 2007; Ziegler et al., 2020). This limitation might be caused by analytic difficulties arising when trying to model interactions between multiple learner characteristics and interventions (Bauer & Shanahan, 2007).

In the present study, we suggest and trial an analytic approach based on latent profile analysis (e.g., Hickendorff et al., 2018) to tackle this issue and model interactions between multiple learner characteristics and instructional variables. As we will discuss and show based on an applied example, this approach allows taking into account multiple learner characteristics and their interactions at the same time. It also bears the potential to overcome problems of low power in contexts involving high-dimensional interactions of learner characteristics, as well as to provide measurement error-corrected estimates and capture non-linearities. We also demonstrate that this approach can provide information of crucial theoretical and practical importance that would be overlooked in more traditional univariate approaches.

We first outline critical issues in the modeling of interactions of learner characteristics with educational interventions. We then outline how a person-centered approach addresses these issues and complement this by an example in which we analyze data about reading instruction in third-grade elementary school classes.

Individualized Instruction

Individualized instruction, the adaptation of instruction to the needs of specific learners or groups of learners, has long been regarded as an important aspect of teaching practice (Corno, 2008; Dockterman, 2018; Tetzlaff et al., 2021). Demographic changes and recently established educational policies, for example, regarding inclusive education, have increased student heterogeneity in classrooms in various countries (Corno, 2008; Decristan et al., 2017; Subban, 2006). Although individualizing instruction has been a topic within educational research for multiple decades (for an overview of early research, see Cronbach & Snow, 1981), these changes have led to newfound interest and a recent increase in the relevance of this topic (Bernacki et al., 2021; Tetzlaff et al., 2022).

The promises of individualization are associated with the impressive effects of one-on-one in-person tutoring (Bloom, 1984; VanLehn, 2011). Endeavors to individualize education can thus be understood as ways to scale up the effects of one-on-one tutoring to larger groups of learners, without having to provide a human tutor

for each learner. This is difficult, especially in regular classrooms where instruction is supposed to address large groups of learners at the same time.

This problem has been addressed by several different approaches, including adaptive teaching (Corno, 2008), differentiated instruction (Constas & Sternberg, 2013), and formative assessment (Deno, 1990). These approaches all have one thing in common: the systematic adaptation of instructional parameters based on the relevant characteristics of specific learners (Tetzlaff et al., 2021). By considering individual learner characteristics such as prior knowledge, cognitive capabilities, and affective/motivational traits and states, teaching agents adapt instructional parameters to achieve optimal fit and maximize potential learning gains across a group of students with heterogeneous preconditions (Grimm et al., 2023).

In order for these instructional adaptations to meet the aim of individualized instruction, instructional parameters need to show differential effectiveness for different learners. Differential effectiveness has been defined by Hunt (1975) as follows: "To consider the differential effectiveness of an educational approach (...) is not simply to point out a few persons to whom the principle does not apply (...). Rather than ask whether one educational approach is generally better than another, one asks, 'Given this kind of person, which of these approaches is more effective for a given objective?'" (Hunt, 1975).

This definition clarifies that when optimizing instruction to the needs of specific learners, it is not sufficient to look into variation in learning outcomes and relate this variation to individual differences in the learners' characteristics. Rather, research into differential effectiveness needs to assign learners to different instructional conditions in a (quasi-) experimental manner to see for whom which instruction works better and what moderates the magnitude (and direction) of differences between conditions. This concept of differential effectiveness can be seen as a variation of the aptitude-treatment interaction (ATI) paradigm (Cronbach, 1957). Without interactions of learner characteristics (aptitudes) and instructional parameters (treatments), some learners would learn better than others, and some instructional approaches would be more effective than others, but adapting the instructional approach to specific learners would have no effect.

The Challenge of Modeling Multivariate Learner Characteristics

The most frequent approach in research on differential effectiveness (or ATIs) is to look at the interaction of one specific learner characteristic with different treatments (e.g., Bracht, 1970; Kalyuga, 2007; Seufert et al., 2009). For example, a well-known effect identified with this approach is expertise reversal: Learners with lower prior knowledge tend to benefit from stronger instructional guidance, whereas, for learners with higher prior knowledge, the same amount of guidance can be unnecessary or distracting (Jiang et al., 2018; Kalyuga, 2007). Similarly, it has been found that learners with lower general reasoning ability can benefit from stronger teacher guidance, while learners with higher general reasoning ability can benefit from stronger self-guidance (Ziegler et al., 2020). Similar interactions have been found between working memory and the effects of conceptual versus fluency activities during

instruction (Fuchs et al., 2014). In most of these applications, the following analytic approach was used. Within a multiple regression approach, main effects of a treatment (e.g., experimental vs. control condition) and a learning prerequisite (e.g., prior knowledge) are modeled. Sometimes, the learning prerequisite is dichotomized beforehand (e.g., via median-split), but this generally leads to a loss in information and is not advisable (Cronbach & Snow, 1981). In addition, an interaction between treatment and learning prerequisite is modeled. If this interaction effect is statistically significant, then researchers would interpret this as support of the hypothesis of an aptitude-treatment interaction/differential effectiveness. This approach, which we call here the traditional univariate approach, has been used in many studies on aptitude-treatment interactions so far (e.g., Coyne et al., 2019; Fuchs et al., 2014, 2019; Vaughn et al., 2019; Ziegler et al., 2020).

These studies have in common that they identified a single learner characteristic that might be of relevance for successful learning in a particular learning context. They then examined interactions of this learner characteristic with different educational interventions. However, as indicated by the practice of teachers adapting their instruction to multiple presumed characteristics of their learners (Corno, 2008), most learning situations likely draw on more than one learner characteristic that determines its effectiveness. Different learner characteristics can interact with one another, leading to different learning outcomes than would be the case for each characteristic on its own (Hooper et al., 2006; Lonigan, Burgess, & Schatschneider, 2018a; Reinhold et al., 2020). This phenomenon has been described under the name of aptitude complexes (Snow, 1987) or trait complexes (Ackerman, 2003). Such multivariate interactions of learner characteristics might occur in many contexts (for theoretical examples, see Cronbach, 1975) and have the potential to provide informative insights for individualized instruction.

An issue that stands in the way of utilizing multivariate learner characteristics to inform instruction is the challenging task of statistically analyzing such interactions. One reason that this is challenging is that, when multiple learner characteristics are taken into account, interaction terms are usually added to regression models to estimate the interactions of these characteristics with instructional parameters. If there are three learner characteristics (e.g., prior knowledge, a learning-relevant cognitive ability, and a motivational variable) that presumably interact in determining the effects of an educational intervention, then all interaction terms up to the fourth order have to be added to the respective model. One problem with interactions of such high order is that they quickly become almost impossible to interpret because, in an almost endlessly complex manner, the interpretation of any effect will always be qualified by another higher-order effect (Cronbach, 1975). A large number of effects are estimated that all depend on each other, leaving researchers with a messy picture about what is going on (Bauer & Shanahan, 2007). Another problem with such approaches is that they require extremely high sample sizes to obtain sufficient statistical power to be informative (Cronbach & Snow, 1981). For example, in a design involving an interaction between a learner characteristic and a factorial intervention variable, easily more than 100 students can be required per intervention condition to yield sufficient statistical power (Cronbach & Snow, 1981). In designs including higher-order interactions, this demand will be even higher,

posing the question of how applied researchers should meet the sample size requirements if their main interest lies in interactions of multivariate learner characteristics with instructional conditions. Finally, interactions of learner characteristics with instructional variables might be of unspecified non-linear shape. For example, the well-known expertise reversal effect describes a curvilinear relation between learners' prior knowledge and the effect of instructional guidance (Kalyuga, 2007). Consequently, the question arises as to how both, interactions between multiple learner characteristics and instruction, as well as non-linear effects, can be best captured in statistical models.

Overall, whereas multivariate learning prerequisites are a topic of great interest for educational researchers, methodological challenges make it difficult to adequately examine such prerequisites and their interactions with educational interventions. Even recent best-practice recommendations for the examination of aptitude-treatment interactions do not tackle these issues (e.g., Preacher & Sterba, 2019).

The Present Approach: Person-Centered Modeling of Learner Characteristics

In the present study, we propose a person-centered approach to the investigation of differential effectiveness involving multivariate learning prerequisites. Person-centered analysis provides a possible avenue of investigation into these processes by moving the focus of analysis from the interactions of single variables to entire persons and their characteristic constellation of learning prerequisites. The specific approach that we propose and use in the present study is that of latent profile analysis (e.g., Hickendorff et al., 2018). In a latent profile analysis, learners are grouped according to their constellations of mean patterns across multiple variables. The different profiles represent groups of learners who systematically differ in their multivariate learner characteristics.

Latent profile analysis is not a novel approach; its foundations, within a framework-labeled latent structure analysis, have been developed during World War II by Henry Lazarsfeld to aid in the selection of military forces (Stouffer et al., 1950). In educational research, a surge of applications can be seen since a contribution by Marsh et al. (2009) illustrated how grouping learners according to their patterns of mean values across multiple variables can provide novel information in comparison to more common factor-analytic models.

Recent research has started applying latent profile analysis and similar methods to model aptitudes. Hooper et al. (2006) identified profiles of learners who differed in their language, problem-solving, attention, and self-monitoring characteristics. Learners with different profiles showed differential development in writing skills during an intervention. With a similar approach, Lonigan, Goodrich, and Farver (2018b) found that language-minority children's profiles of proficiency in their first and second language predicted their development of early literacy skills in preschool. Reinhold et al. (2020) identified subgroups of sixth graders with different engagement profiles that were systematically related to their development in mathematics

achievement. Finally, Grimm et al. (2023) applied latent profile analysis on outcome variables of an instructional sequence to model the differential effects of experimental conditions on third graders' development of multiple reasoning skills.

In the present approach, we extend these applications of latent profile analysis in order to tackle the described issues in research on the differential effectiveness of instruction. Specifically, one characteristic that the above-mentioned studies have in common is that, while they did investigate how learner profiles representing multivariate learner characteristics relate to learning outcomes, they did not investigate how these profiles interacted with the effects of specific instructional parameters. Only when we understand such interactions, we can speak of differential effectiveness and use the data as a foundation for providing different treatments to different learners (Hunt, 1975). Therefore, in the present study, we suggest an extension of this approach in which latent profile analysis is used in a two-step procedure for examining the differential effectiveness of educational interventions.

In the first step, learner characteristics that presumably interact with the effects of instructional parameters are grouped according to latent profile analysis. In principle, any other clustering method could be used to achieve this goal, such as cluster analysis. We bring forward latent profile analysis because it is a model-based clustering approach, bringing with it advantages such as the possibility to obtain model estimates that are corrected for measurement error (Hickendorff et al., 2018). In the second step, learner profiles are related to the effects of instructional parameters to examine how the multivariate learner profiles interact with educational interventions. This can be achieved by different approaches (see, e.g., Nylund-Gibson et al., 2014). We propose using the manual BCH approach in research on differential effectiveness (Gudicha & Vermunt, 2013; Nylund-Gibson et al., 2019). This approach enables relating latent profiles to any parameters that can be modeled in the framework of structural equation modeling, correcting model estimates for measurement error.

By grouping learners into profiles using this method and then comparing the effects of educational interventions between those groups, their differential effectiveness across different patterns of multivariate learning prerequisites can be examined. Overall, this approach has the advantage of being able to group many learners into just a few distinct categories. This feature helps to achieve high statistical power while capturing multivariate and potentially non-linear information across multiple learning prerequisites in a much more concise manner than a regression analysis with higher-order interactions (Bauer & Shanahan, 2007).

We are not aware of prior research taking this person-centered approach to investigating differential effectiveness as we define it here. There are, however, studies finding differential effectiveness of treatments for subgroups of learners that have been grouped with other approaches. One example is a study by Hofer et al. (2018) in which intelligence and gender interacted in their effect on the efficacy of cognitively activating instruction in physics. In the present study, we use latent profile analysis to examine the differential effectiveness of different reading interventions for different groups of learners who systematically differ in their prerequisite skills for reading according to the well-known model of the simple view of reading.

Multivariate Learning Prerequisites for Reading—the Simple View of Reading

Reading comprehension is a complex skill that is constituted by an interplay of several different components (Kendeou et al., 2016). One of the most prominent theories to explain how these components relate and how they interact to form the construct of reading comprehension is the simple view of reading (SVR; Hoover & Gough, 1990). According to that theory, reading comprehension can be modeled as the product of language comprehension (LC) and decoding (D) abilities.

Decoding is defined as the ability to recognize printed words accurately and quickly. It covers both the fluency as well as the accuracy of the grapheme-phoneme conversion process. Language comprehension is the ability to extract and construct literal and inferred meaning from linguistic discourse represented in speech and can include knowledge about phonology, syntax, or semantics as well as background knowledge or inferential skills (see Hoover & Tunmer, 2022).

The multiplicative nature of the model implies a difficulty in compensating for deficits in one of the two abilities with improved performance in the other. Evidence of this relation has been found in multiple alphabetic languages (see Hjetland et al., 2020 for a comprehensive review). This two-dimensional conceptualization means that readers fall into one of four quadrants of reading comprehension: *Good readers* have good word-reading/decoding skills, complemented by good comprehension abilities, *poor readers* lack in both abilities, while *poor decoders* as well as *poor comprehenders* show good performance in one of the two components, combined with poor performance in the other.

Since decoding and language comprehension skills have reliably been identified as distinct predictors of reading comprehension across multiple alphabetic languages (Ehm et al., 2023; Hjetland et al., 2020; Lonigan, Burgess, & Schatschneider, 2018a), we used this distinction as a starting point for our latent profile analysis. For task selection, we orientated on recent SVR publications (Hoover & Tunmer, 2018, 2022) but also on a small body of research on reading profiles in elementary school children (Foorman et al., 2017; Tambyraja et al., 2015). Foorman et al. (2017), for example, used a word recognition task (decoding), a vocabulary task (semantic knowledge), and a syntactic knowledge task to identify reading profiles in third- to tenth-grade students.

We refer to the SVR for this study because it serves as a good basis for considering both theoretically grounded multivariate aptitude profiles, as well as their interaction with instruction. While we do acknowledge that the SVR does not fully describe all possible aspects of reading performance (see Castles et al., 2018 for a discussion), it is still widely used in current research on reading and is able to explain substantial amounts of variance in reading comprehension and its development (e.g., Lervåg et al., 2018; Lonigan, Goodrich, & Farver, 2018b). For the purpose of the present study, it is important to note that the SVR implies specific predictions regarding optimal reading instruction for learners with different preconditions. A straightforward deduction would be that poor comprehenders benefit from instruction that specifically targets comprehension, while poor

decoders would benefit from instruction that specifically targets decoding. This is supported by a meta-analysis by Galuschka et al. (2014) showing that, in general, interventions that target children's specific deficits are more effective in alleviating their reading difficulties than more general approaches. In the present study, we examine how the instructional emphasis that teachers put on different aspects of reading instruction meets the actual needs of students with these different constellations of learner characteristics and thereby helps them improve their reading comprehension.

Instructional Foci in Third-Grade Reading Lessons

In the present study, we examine the differential effectiveness of different foci in reading instruction in third-grade classes within the context of German elementary schooling. The curriculum for third-grade students in the federal German state of Hesse (the location of our study) generally puts a strong emphasis on strengthening students' reading motivation, teaching advanced reading abilities such as passage comprehension and summarizing texts, and fostering vocabulary acquisition—for example, by finding synonyms of words or using previously unknown words in exemplary sentences (Hessian Ministry of Education, 2021). Which of these aspects is emphasized at which point in time, and how much time is invested in each of these aspects, will naturally vary from teacher to teacher. We hypothesize that the instructional foci teachers choose will differentially affect students based on their individual learning prerequisites. While these instructional foci are not identical to specific training of these reading-related abilities, we still assume that it is useful to refer to findings from related training studies to inform our hypotheses about their differential effectiveness.

Fostering reading motivation can take many forms, for example, encouraging learners to seek out and read literature based on their own interests. Reading motivation is positively related to reading comprehension (Kuşdemir & Bulut, 2018), and longitudinal studies indicate a reciprocal relation between reading comprehension and reading motivation in second and third grades (e.g., Schiefele et al., 2016). This strongly implies that, at least for a specific subset of students—those who possess the necessary skills to read texts without instructional support—fostering reading motivation could lead to increased reading comprehension over time. An assumed mediating mechanism of this relationship is the increased frequency of reading in out-of-school contexts in highly motivated readers (Guthrie et al., 1999). In this case, it is likely that the positive effects of interventions focusing on reading motivation appear with a considerable delay.

Advanced reading abilities comprise several different techniques dealing with passage comprehension, such as highlighting important aspects, writing short summary sentences for specific passages, and rephrasing content in one's own words. As advanced reading abilities can be addressed by a broad spectrum of instructional approaches, it is difficult to derive specific predictions from the training literature. It is reasonable to assume that a focus on comprehension skills mainly benefits those

children who specifically struggle with comprehension, as opposed to those struggling with decoding, or with both.

While vocabulary training has repeatedly been shown to increase vocabulary size or quality (e.g., Segers & Verhoeven, 2003), transfer to reading comprehension abilities seems to be limited (Mezynski, 1983). This is especially interesting given the strong correlation between vocabulary size and reading comprehension (Carroll, 1993; Freebody & Anderson, 1983). These results are not completely unambiguous though. A meta-analysis of these transfer effects by (Elleman et al., 2009) found significant variability between studies, mostly dependent on the type of vocabulary measure used, but also in relation to students' learning prerequisites. This variability might point to the efficacy of vocabulary training for learners with specific learning prerequisites, but not for others. Therefore, investigating the differential effectiveness of that transfer is of particular interest: Are there children who make progress in reading comprehension as a result of vocabulary training, and how are their learning prerequisites constituted?

The Present Study

We employ the person-centered approach of latent profile analysis to examine the differential effectiveness of different instructional foci for third graders' development of reading comprehension. Our intention behind this study is twofold. First, we explore the methodological approach outlined above. Second, although exploring our methodological approach is the main aim of this study, we do this within the context of reading instruction. Therefore, we will also explore and discuss implications for adaptive reading instruction that can be drawn from our findings. Specifically, we pose the following three research questions:

- 1) Which latent profiles of reading abilities exist within a group of third-grade learners?

We employed latent profile analysis to examine which profiles regarding decoding and comprehension abilities exist among the learners. Learner profiles will be based on four variables that describe their prior knowledge in reading comprehension and their levels on three learning prerequisites related to the simple view of reading, namely, decoding, syntactic comprehension, and vocabulary knowledge. Although this research question is not yet related to our main research question—the potential of the proposed analytic approach—extracting profiles and embedding these in available theory are prerequisite first steps for examining and interpreting the differential effectiveness of instruction across learners within different profiles.

- 2) Do different instructional foci show differential effectiveness across readers with different profiles over the course of one school year?

To examine this question, we assessed teachers' instructional foci throughout the course of one school year. By an instructional focus, we refer to the emphasis that they put on different aspects of reading instruction (i.e., vocabulary, advanced reading abilities, and reading motivation). Based on the rationales outlined above, we hypothesized that a focus on reading motivation will benefit good readers—

those who already show good decoding and word-reading skills. Furthermore, based on the simple view of reading, we expected that learners with poor comprehension skills would benefit from advanced reading training. This being an exploratory study, we had no specific hypothesis about which learners would particularly benefit from vocabulary training.

- 3) Does our analytic approach reveal results that would not be apparent from, or even contradictory to a more traditional multiple regression approach that considers one learning prerequisite at a time?

We will conduct the above-described typical multiple regression model for each of the four learning prerequisites (reading comprehension, decoding, syntactic comprehension, vocabulary) with one variable at a time. We will then compare findings from this common approach with our newly proposed approach to examine whether our approach has the potential to reveal additional insights.

Method

The current study uses data from a larger research project running from 2018–2020, which was approved by the Ethics Committee of BLINDED. Data were collected in two cohorts, one in the school year 2018/2019 in the state of Hesse in Germany and the other in the school year 2019/2020 in the states of Hesse and Lower Saxony. Each cohort completed a pretest at the beginning of the school year and a posttest before the summer break. All tests were administered by research assistants, apart from the posttests in the school year 2019/2020 that had to be administered by the respective teachers due to pandemic-related school lockdowns. Since the tests were easy to administer, we do not expect that this caused a significant reduction in data quality. The teachers and their students participated in the study on a voluntary basis and did not receive any compensation. As the recruitment was done in cooperation with the Ministry of Education of Hesse, we did not put an upper limit on the sample size. Available simulation studies indicate that a sample size of between 100 and 500 is needed to identify the correct number of profiles in a latent profile analysis (Edelsbrunner et al., 2023; Nylund et al., 2007). Data were analyzed using R, version 4.0.2 (R Core Team, 2021) and the package MplusAutomation, version 0.78-3 (Hallquist & Wiley, 2018) as well as Mplus 8 (Muthén & Muthén, 2017). As this is an exploratory study, its design and its analysis were not pre-registered.

Sample

We relied on two different samples for the analysis: The whole sample was used for the analysis related to research question one, while a subsample was used for the analysis of research questions two and three. For the whole sample, the teacher group consisted of 59 teachers from 36 schools in Hesse and 16 teachers from 3 schools in Lower Saxony. Teachers were, on average, 41 (SD = 9.19) years old. Two of them were male, and the remaining 73 were female. They reported an average of 13.26 (SD = 7.58, range = 4–39) years of general teaching experience and 5.68 (SD

= 5.56, range = 0–27) years of experience teaching third-grade classes. The teachers nominated some of their students ($N = 517$ in total) to participate in individual testing sessions. Teachers could select up to eight children. They were asked to prioritize children with reading difficulties and then add children to be representative of the class, based on their own criteria. These students were on average 8.32 (SD = 0.56) years old and 48% speak German as their first language, 23% grew up bilingual, and 30% speak German as their second language. We used the data of these nominated students from all classes for the first set of analyses, with the aim of identifying latent profiles of readers.

For the analysis related to the second and third research questions, which dealt with the differential effects of instructional foci, we used a subsample of 52 teachers in Hesse who completed additional online questionnaires on their instructional foci. Teachers in that sample were on average 38.04 (SD = 8.19) years old, with 11.91 (SD = 6.41) years of teaching experience, 4.60 (SD = 4.34) of which in third-grade classes. The respective student sample comprised 217 students that were on average 8.34 (SD = 0.56) years old, with 59% speaking German as their first language, 19% growing up bilingual, and 22% speaking German as their second language.

Assessment Materials

Teachers' Instructional Foci

Throughout the school year, teachers were presented with an online questionnaire every 3 weeks in which they were asked about their teaching practices during that time. They were asked to fill out this questionnaire 8 times during the school year, and the average participation rate was 4.10 times (range 1 to 8). Teachers' self-reported practices in reading instruction showed a moderate to high stability (ICC(1) = 0.21–0.51) across assessments, indicating that (a) they did not vary their instructional focus much and (b) the aggregated measure can be seen as a reliable estimate of the instructional landscape over the whole school year. The teachers with only one rating did not enter into the calculation of the ICCs. Fleiss (1986) proposes that an ICC(1) > 0.15 serves as a reasonable benchmark for these kinds of assessments, while LeBreton and Senter (2008) recommend using ICC(1) values of .01, .10, and .25 as benchmarks for small, medium, and large effects (of group membership on individual rating). To examine the instructional foci of interest in this study, we asked them how much emphasis they put on “vocabulary training,” “advanced reading abilities (sentence and passage comprehension),” and “reading motivation.” The other measures in the questionnaire pertained to the organization of instruction (e.g., peer-teaching, individualized attention) or were not reflective of regular third-grade reading instruction (e.g., a focus on precursor abilities). All of these were answered on a 4-point Likert scale with the anchors *never* and *often*. Several previous studies have shown that teacher self-report ratings correlate with student and observer ratings of the same construct (Clunies-Ross et al., 2008; Fauth et al., 2014), especially when asked retroactively about concrete behavior in the recent past (Tetzlaff et al., 2022). For the current analyses, we used average values across time for each of the

three selected instructional foci. By regularly asking teachers which of these aspects they emphasized in their classes and aggregating these measures over the whole school year, we can create a picture of the instructional landscape in that specific classroom.

Decoding Ability

As our measure of decoding ability, we used the pseudoword part of the Salzburg Reading and Writing Test SLRT II (Moll & Landerl, 2010). In this test, children are asked to read a written list of pseudowords aloud while the experimenter keeps track of the amount of correctly read words. We did not estimate the internal consistency of this measure in our sample as this would have required recording the full sessions with the participants, which we did not do. The test, however, generally shows very high reliability estimates, and it is a test commonly used for diagnostics of individual children that require high precision. The manual reports the reliability (measured via parallel tests) as between .90 and .98 (Moll & Landerl, 2010). For statistical analysis, we used the amount of correctly read pseudowords within 1 min.

Syntactic Comprehension

As our measure of syntactic comprehension, we used the screening in the TSVK (Sieg Müller et al., 2011). In this test, children are asked to select the one picture, out of a set of three, that corresponds most to a sentence that was read aloud to them. The items in the test showed a satisfactory internal consistency ($\alpha = 0.66$, $\omega = 0.69$). For statistical analysis, we used the amount of correctly selected pictures as a measure of learners' syntactic comprehension, which we used as an indicator variable for their language comprehension.

Vocabulary Knowledge

As our measure of vocabulary, we used a measure of productive vocabulary (WWT; Glück, 2011). In this test, children are presented with pictures on a computer screen and then are asked to name the depicted concept. The test consists of 40 nouns, verbs, or adjectives and provides a list of synonyms/alternative answers that would be counted as correct. The items in the test showed a high internal consistency ($\alpha = 0.88$, $\omega = 0.91$). For statistical analysis, we used the sum score of correct responses as a measure of learners' vocabulary knowledge. Vocabulary knowledge was operationalized as one of two indicators of the language comprehension component. The tests for vocabulary, syntactic comprehension, and decoding were administered by trained research assistants in individual sessions.

Reading Comprehension

As our measure of general reading comprehension, we used the pen-and-paper version of the ELFE II (Lenhard & Schneider, 2006). The ELFE II measures reading

comprehension at the word, sentence, and text level. For the word comprehension task, children are presented with a picture and a group of four written words and asked to select the word that matches the picture. For the sentence task, children are presented with an incomplete sentence and asked to select one of five written words to complete it. For the text task, children are asked to read short passages and then select the statement that best corresponds to the passage out of a choice of four. The items in the test showed a high internal consistency ($\alpha = 0.96$, $\omega = 0.97$). The ELFE was administered once at the beginning (T1) and once at the end (T2) of the school year as a pen-and-paper version simultaneously to the whole class. For the first cohort, as well as the pretest of the second cohort, the tests were conducted by our trained research assistants; the posttest of the second cohort was conducted by the teachers themselves instead. For statistical analysis, we used the mean score of each student across the different sub-skills measured by the ELFE as a general indicator of reading comprehension.

Analytic Approach

In order to estimate student profiles of reading skills and reading comprehension, we conducted latent profile analyses (Ferguson et al., 2020; Harring & Hodis, 2016; Hickendorff et al., 2018) using the software package Mplus 8.3 (Muthén & Muthén, 2017). Before undertaking these analyses, we z -standardized the indicator variables for improved interpretability.

The indicator variables used as the basis of the student profiles were learners' scores on decoding (SLRT), language comprehension (two scores: one each from the TSVK and the WWT), and reading comprehension (ELFE at T1).

The logic behind including the T1 reading comprehension measure in the profile analysis is based on three reasons. First, we needed to control for prior reading comprehension in order to model baseline-corrected reading comprehension at the end of the academic year. Not controlling for such an important prior knowledge measure would likely substantially decrease statistical power. Second, by including the pretest measure within the profiles, we could capture potentially non-linear interactions between this measure and the further indicator variables. Third, from a theoretical perspective, it is useful to include the theoretical outcome of latent profiles directly in the profile analysis, since this very likely improves statistical power and precision in extracting the profiles that are theoretically related to this measure.

Based on these four indicator variables, in a stepwise manner, we increased the number of profiles from one to seven, after which it was evident that fit indices were getting worse and model convergence was not possible anymore. As is common practice in latent profile analysis, the model with the actual number of profiles interpreted and used for further analyses was then selected based on fit indices and theoretical considerations (Ferguson et al., 2020; Harring & Hodis, 2016; Hickendorff et al., 2018). To this end, we relied in particular on the fit indices BIC, aBIC, and the VLMR-likelihood ratio test (Edelsbrunner et al., 2023; Ferguson et al., 2020; Hickendorff et al., 2018; Lo et al., 2001). For the BIC and aBIC, lower estimates indicate

a better relative model fit (for explanations of these fit indices, see Edelsbrunner et al., 2023), and for the VLMR, the model with the highest number of profiles reaching significance should be selected (Ferguson et al., 2020; Harring & Hodis, 2016; Hickendorff et al., 2018).

Regarding technical specifications, each model was estimated with 400 initial random starts, of which the most promising 100 were used for further estimation. The estimation method was maximum likelihood with expectation-maximization optimization and Huber-White standard errors that are robust against multivariate kurtosis and heteroscedasticity (Freedman, 2006). We took the multilevel structure of the data into account through a cluster-robust estimation of the standard errors in all modeling steps (Muthén & Muthén, 2017). In research in which research questions do not require separating variance on multiple levels, this is the preferred method for accommodating multilevel structure in the data (McNeish et al., 2017). For convergence criteria during estimation, we used the Mplus defaults, and we accepted a model as converged when the best likelihood was achieved multiple times. Apart from the means, we also estimated the variances freely within each profile (Edelsbrunner et al., 2023; Ferguson et al., 2020; Hickendorff et al., 2018).

To examine the second research question, concerning the differential effectiveness of instructional foci, we related profile membership to the reading comprehension of the students in the subsample at the end of the school year and to teachers' foci in reading instruction. To this end, we decided on the BCH method (for details on this method and its implementation, see Asparouhov & Muthén, 2014). This approach has an advantage in that it allows for a modeling of differential effectiveness for students in different profiles while correcting for measurement error, in a way similar to a structural equation model (Vermunt, 2010). We regressed the participants' reading comprehension at the posttest on the instructional foci of their teachers. We used reading comprehension as the sole outcome measure because of our applied focus on how reading comprehension can be fostered in school. In a last addition, to check for the interaction between students' reading profiles and teachers' instructional foci, we defined derived parameters. These parameters indicated differences between the different reading profiles in the regression weights of reading comprehension on teachers' instructional foci. These parameters allowed us to check for differences between profiles, as well as to identify the main effects of each instructional focus across all profiles.

For statistical inference, we present and interpret 90% confidence intervals for all focal model parameters. The present analyses follow a more exploratory than confirmatory approach, which undermines the reliability of significance testing. Instead, we focus on confidence intervals and interpret them as follows. If a confidence interval excludes 0, we interpret this as evidence pointing toward a hypothesis that should be further investigated in future research. If a confidence interval includes 0, we cautiously interpret this as a lack of evidence for an effect of interest. Since analyses of aptitude-treatment interactions generally require very large samples (Cronbach & Snow, 1981), we chose 90% intervals to ensure a good balance between alpha and beta errors. For interested readers, the results with 95% confidence intervals are provided in the analysis outputs.

To examine the third research question, we conducted a separate data analysis, employing a univariate approach. We entered all the indicator variables (vocabulary, syntactic, decoding, and reading) as well as their interactions with teachers' instructional foci into separate multiple regression models (one for each indicator), predicting learners' reading comprehension at posttest. We also conducted a more complex multiple regression with all predictors in the same model. These analyses were conducted using the lme4 package (Bates et al., 2015) in R, including a random intercept across school classes to cover multilevel structure in accordance with typical ATI analyses (e.g., Coyne et al., 2019; Fuchs et al., 2019; Vaughn et al., 2019). All data, syntaxes, and model-output files are available under https://osf.io/a97gv/?view_only=73f3249ac61f4b618415d3116e1b164f.

Results

Descriptive Statistics

Students in our sample achieved slightly below-average scores across all reading abilities at the pretest, when compared with a norm sample (*t*-values ranging from 39.97 for vocabulary to 46.61 for reading comprehension). This is probably due to the teacher-selected student sample: Teachers were asked to prioritize children with reading difficulties and then add children to be representative of the class. This leads to a slight over-sampling of struggling readers. The negative correlations between indicators of decoding and language comprehension (see Table 1) have also been reported for samples low on reading comprehension (Hoover & Gough, 1990).

The means of the aggregated teacher self-reports on their instructional foci were consistently in the upper half of the 4-point Likert scale (3.0 for vocabulary training and 3.23 for advanced reading abilities and reading motivation) indicating a general tendency to report the presence, rather than the absence, of specific instructional foci. This is consistent with findings concerning biases resulting from social or educational desirability of certain types of instruction (Kopcha & Sullivan, 2007).

Table 1 Correlations of the indicator variables

	Decoding (SLRT)	Syntactic comprehension (TSVK)	Vocabulary (WWT)	Reading comprehension (ELFE)
Decoding (SLRT)	–	–.10	–.23	.62
Syntactic comprehension (TSVK)		–	.61	.17
Vocabulary (WWT)			–	.11
Reading comprehension (ELFE)				–

Student Profiles of Learning Prerequisites

To investigate the first research question, the first step was to decide how many separate profiles were present in the data. Figure 1 depicts the different fit indices for the models with different numbers of profiles. As shown in Fig. 1, the BIC had its lowest value with six profiles and the aBIC with seven profiles. The VLMR test, however, indicated significant improvement in model fit only with up to four profiles ($p = .02$) but not, for example, with five profiles ($p = .07$). In addition, the BIC and aBIC showed a visible decrease in strength of improvement from the four- to five-profile models. Given these indications and the straightforward interpretability of the four-profile solution, we decided to proceed with the four-profile model. Results with the five-profile model can be found in the supplementary materials.

The four identified profiles can be labeled in accordance with the SVR (see Fig. 2). The poor readers ($n = 78$) are mainly characterized by their low performance in the reading comprehension task, but they also score well below the mean on all the other indicator variables. The good readers ($n = 88$) have extraordinarily high performance in reading comprehension, complemented by scores well above the mean on all the other indicator variables. The poor comprehenders ($n = 195$) have slightly above-average reading comprehension skills, strong decoding skills, weak syntactic comprehension, and below-average vocabulary. The poor decoders ($n = 156$) have slightly below-average reading comprehension skills as well as decoding skills, balanced by strong performance in the language comprehension indicators. The profiles of poor readers and poor comprehenders both had a significantly higher than average amount of students with German as the second language. An overview

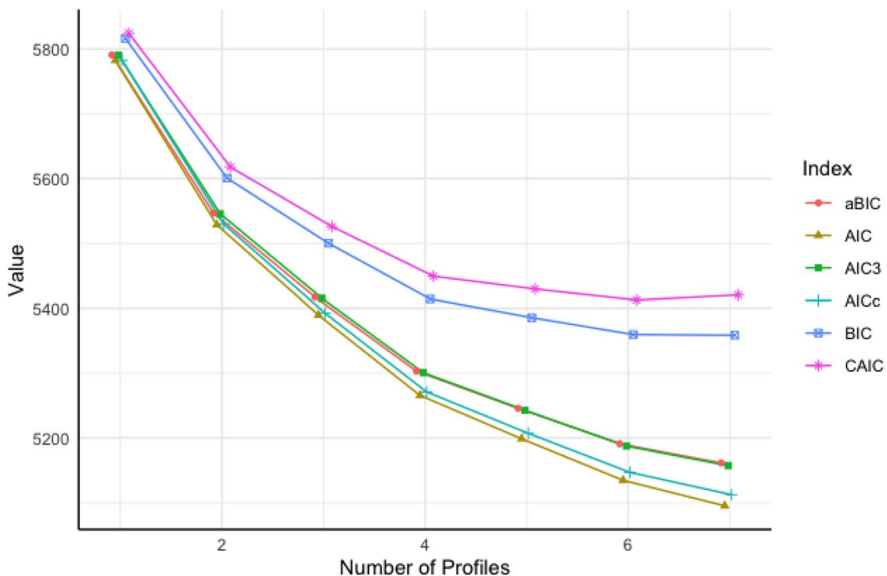


Fig. 1 Fit indices for different numbers of profiles. Note: the elbows in the curves indicate declines in information gained from the addition of more profiles

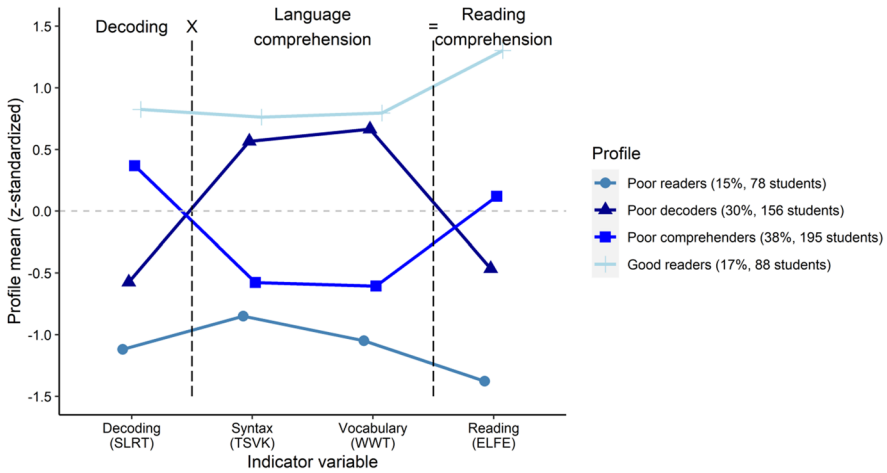


Fig. 2 Profiles of readers, labeled according to the simple view of reading

of the language status of students in different profiles can be found in the supplementary materials (Table S2).

In sum, by employing latent profile analysis, we were able to identify informative profiles across multiple relevant learner characteristics. The strong correspondence of these profiles with the SVR speaks to their validity and provides some indications concerning potential interaction with instruction, which we investigated next.

Differential Effectiveness of Instructional Foci for Students in Different Profiles

For the second research question, concerning the differential effectiveness of instructional foci for students in the different profiles, we used data from the subsample of third graders whose teachers ($n = 52$) had participated in the surveys assessing their instructional foci. Of the $n = 217$ students eligible for this analysis, 89 (41.01%) belonged with the highest probability to the poor comprehenders profile, 66 (30.42%) to the poor decoders, 46 (21.20%) to the good readers, and 16 (7.37%) to the poor readers. Please note that, since only a few students belonged to the poor reader profile, we were not able to investigate differential effectiveness for those students due to too little statistical information.

We first extracted the information on students' profiles from the latent profile analysis conducted for the first research question. We then used the bias-correcting BCH method (Asparouhov & Muthén, 2014) to set up multiple regression models predicting reading comprehension at the end of the school year separately for each profile from teachers' instructional foci. Figure 3 depicts the results from this model. The standardized regression weights (y -axis) are shown for each different instructional focus (x -axis) and each different profile (symbol). A detailed list of the regression weights and their respective standard errors can be found in the supplementary materials (Table S1). When looking at the profile-specific regression weights (Fig. 3; Table S1), we can see that—consistent with the

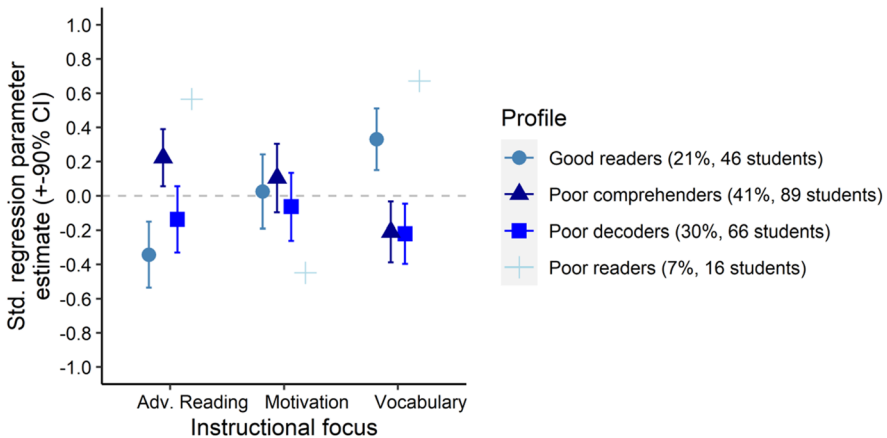


Fig. 3 Parameter plot of regression weights describing effects of instructional foci on reading comprehension for students in different profiles

predictions made by the SVR—poor decoders do not improve when vocabulary is the focus. Poor comprehenders also do not improve with a focus on vocabulary, but they benefit from a focus on advanced reading skills. Good readers benefit from a focus on vocabulary and suffer when the focus is on advanced reading skills. Since the group of poor readers which also had self-report data from their teachers only consisted of approximately 16 students, we are unable to make any reliable statement about them.

Direct evidence for differential effectiveness is present if the different instructional foci show varying effect sizes for learners in the different profiles. To examine this statistically, we set up contrasts that represented group differences in the effects of the instructional foci for learners in the different profiles. The first information that we examined before inspecting differential effectiveness was the simple main effects of teachers' foci on students' reading comprehension at the end of the school year. These effects, estimated in a simple multiple regression, were all close to zero (Table 2).

While none of the instructional foci showed any main effect for all participants, the specific contrasts for the different instructional foci (see Table 2) show strong differential effectiveness of vocabulary training between good readers and poor decoders, as well as between good readers and poor comprehenders. They also indicate a strong differential effectiveness of training advanced reading abilities between good readers and poor comprehenders, as well as between poor decoders and poor comprehenders. For the fostering of reading motivation, no meaningful differential effectiveness could be found.

To better illustrate the magnitude of these effects, Fig. 4 shows the projected reading comprehension scores at posttest (*Y*-axis) for learners in the four different profiles (different symbols/colors), being instructed by teachers who put little, average, or high emphasis (*X*-axis) on either of the three instructional foci.

Table 2 Main effects and specific contrasts on vocabulary, advanced reading abilities, and motivation, including 90% confidence intervals

Contrast	Beta	SE	Lower 5%	Upper 5%
Advanced reading focus				
Main effect	0.018	0.116	-0.2	0.191
PD vs. PC	0.361	0.134	0.14	0.581
GR vs. PC	-0.567	0.15	-0.813	-0.32
GR vs. PD	0.206	0.163	-0.063	0.457
Motivation focus				
Main effect	-0.005	0.119	-0.172	0.209
PD vs. PC	0.168	0.17	-0.112	0.448
GR vs. PC	-0.079	0.166	-0.353	0.194
GR vs. PD	-0.089	0.971	-0.346	0.168
Vocabulary focus				
Main Effect	-0.034	0.095	-0.19	0.122
PD vs. PC	0.012	0.166	-0.261	0.284
GR vs. PC	0.54	0.164	0.272	0.81
GR vs. PD	0.552	0.167	0.278	0.827

Poor readers were excluded due to their limited sample size
PD poor decoders, *PC* poor comprehenders, *GR* good readers

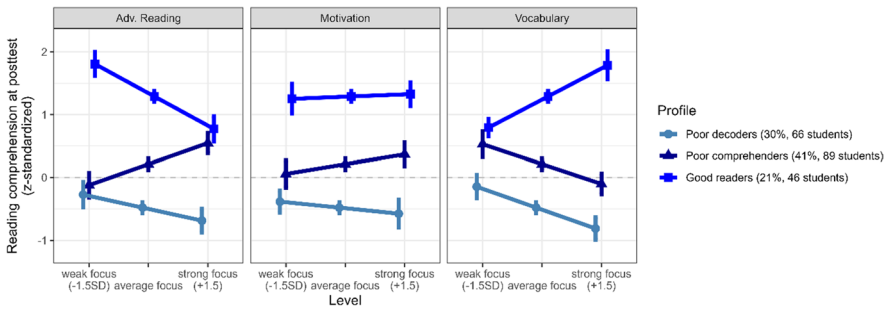


Fig. 4 Model-predicted reading comprehension at posttest, given a specific profile and instructional focus

Comparison to Results from the Multiple Regression Approach

For a better comparison, we contrast it with a “traditional” variable-centered approach in which multiple regression models with 2-way interactions between each instructional focus and each learning prerequisite are employed. To acknowledge the nested structure of the data, we added a random intercept for the different school classes (as is the current best practice as shown by Coyne et al., 2019; Fuchs et al., 2019; and Vaughn et al., 2019). The results of this analysis can be found in Table 3.

As can be seen in Table 3, some of the conclusions that can be drawn from these analyses differ from those that can be drawn from the person-centered approach, especially when the pronounced differential effectiveness of a focus on vocabulary that was apparent in the person-centered approach was not replicated in these

Table 3 Results of the variable-centered multiple regression analyses, including 90% confidence intervals

	Beta	SE	Lower 5%	Upper 5%
Model 1—Vocabulary				
Vocabulary focus	0.35	1.45	− 1.95	2.63
Advanced reading focus	− 0.71	1.55	− 3.15	1.74
Motivation focus	1.12	1.66	− 1.49	3.74
Vocabulary knowledge	0.55	0.72	− 0.58	1.79
Reading comprehension (pretest)	14.57	0.74	13.40	15.86
Vocabulary focus *Vocabulary knowledge	0.43	1.29	− 1.66	2.50
Advanced reading focus *Vocabulary knowledge	− 2.20	1.19	− 4.13	− 0.29
Motivation focus *Vocabulary knowledge	2.55	1.29	0.51	4.71
Model 2—Decoding				
Vocabulary focus	0.07	1.38	− 2.12	2.25
Motivation focus	3.02	1.67	0.38	5.72
Advanced reading focus	− 2.32	1.55	− 4.80	0.14
Reading comprehension (pretest)	13.49	0.85	12.15	14.94
Vocabulary focus *Decoding	− 2.18	0.86	− 3.65	− 0.82
Advanced reading focus *Decoding	3.06	1.67	1.44	4.87
Motivation focus *Decoding	− 2.48	0.93	− 4.03	− 1.00
Model 3—Syntactic comprehension				
Vocabulary focus	− 0.24	1.48	− 2.58	2.10
Advanced reading focus	0.73	1.59	− 1.79	3.27
Motivation focus	0.11	1.65	− 2.50	2.72
Reading comprehension (pretest)	19.24	0.96	17.72	20.87
Syntactic comprehension	1.73	0.75	0.54	2.98
Vocabulary focus* Syntactic comprehension	0.10	1.18	− 1.85	1.99
Advanced reading focus* Syntactic comprehension	− 2.24	1.19	− 4.20	− 0.33
Motivation focus* Syntactic comprehension	1.83	1.24	− 0.13	3.95
MODEL 4—Reading comprehension				
Reading comprehension (pretest)	14.07	0.69	12.98	15.26
Vocabulary focus	0.71	1.48	− 1.63	3.04
Advanced reading focus	− 2.34	1.74	− 5.11	0.41
Motivation focus	2.31	1.82	− 0.57	0.41
Vocabulary focus* Reading comprehension	− 2.08	0.94	− 3.66	− 0.59
Advanced reading focus* Reading comprehension	1.92	1.08	0.22	3.77
Motivation focus* Reading comprehension	− 1.58	1.10	− 3.35	0.22

variable-centered analyses, with only negative interactions between it and different learning prerequisites being obtained. Instead, some differential effectiveness of a

focus on reading motivation can be found: motivation-oriented teaching appears to be especially effective for students with good vocabulary and especially ineffective for students with good decoding skills.

Most importantly, these analyses do not really inform the choice of treatment for a specific child as the differential effectiveness of instructional foci can be vastly different, depending on the specific variable that was investigated. We therefore set up a second, more complex regression model in which all predictors and 2-way interactions between learning prerequisites and instructional foci were added concurrently (Table S1). The results from this analysis only showed a small positive interaction between decoding skills and a focus on advanced reading abilities, which is consistent with both the person-centered approach and the separate variable-centered models. The positive interaction between a focus on reading motivation and the vocabulary knowledge of students as well as the negative interaction between this focus and decoding that were visible in the separate multiple regression models (Table 3) were not visible in the concurrent model (Table S1) or the person-centered approach (Table 2).

Overall, neither in the separate multiple regression models nor in the model incorporating all variables concurrently, the full results pattern from the person-centered approach were visible.

Discussion

In the present study, we used a person-centered approach to modeling multivariate learning prerequisites and their interactions with instructional parameters. We ran a latent profile analysis to identify learner profiles based on several sub-skills of reading and then analyzed which instructional foci proved effective for each learner profile. Patterns of differential effectiveness emerged despite a lack of main (average) effects of the different instructional foci across all students, indicating that none of the foci is, by itself, preferable to the others when applied across all students. Only by taking the specific learning prerequisites of students into account can an informed selection of instruction be made. We interpret these results to have implications for educational research in general and for reading instruction in particular.

The Benefits of Person-Centered Analysis for Research on Differential Effectiveness

How to model multivariate learner characteristics and their interaction with instruction has remained a more or less unresolved problem in educational psychology (Cronbach, 1975). In this study, we make a strong case for utilizing person-centered analysis to group learners in accordance with their pattern of means across multiple relevant characteristics as well as to investigate the differential effectiveness of specific instructional parameters for these groups. Instead of asking for which *levels* of a certain characteristic a specific treatment is most effective, researchers should ask for which *learners*—with their specific constellation of learning prerequisites—a

treatment is most effective. Person-centered analysis such as latent profile or class analysis allows researchers to pose and answer this question. This approach has the added benefit of also identifying non-linear relations between aptitudes and treatment effectiveness that would be lost, or at least be difficult to track and interpret, in traditional variable-centered analyses (Bauer & Shanahan, 2007). Specifically in the domain of reading, non-linear effects are increasingly used to explain developmental trajectories as well as learning gains in reading comprehension (e.g., Ehm et al., 2023; Hjetland et al., 2019).

Utilizing the 3-step BCH method (Asparouhov & Muthén, 2014) further allows for the integration of the class-specific regression terms without either biasing the profile estimation or embezzling measurement error. In sum, person-centered analysis circumvents several problems associated with modeling multivariate learner characteristics and their interaction with instruction and thus can be seen as a promising approach for future studies in these domains. With its multivariate nature, our approach goes beyond the state-of-the-art methods for the analysis of differential effectiveness/aptitude-treatment interactions.

Implications for Reading Instruction

Besides the main aim of demonstrating how fruitful the application of person-centered analysis can be for investigating differential effectiveness, our results also provide some support for the simple view of reading (Hoover & Gough, 1990). The four identified profiles suggest that the SVR is valid for describing reading performance on a level that is useful for classroom instruction. This is not a completely new finding: Torppa et al. (2007) used latent profile analysis to identify five subgroups of reading performance that correspond to the four profiles predicted by the SVR, with an additional profile of average readers.

Wolff (2010) identified eight latent profiles across ten reading-related abilities. Of these eight profiles, three proved to be especially stable—good readers, poor decoders (dyslexics), and poor comprehenders. Foorman et al. (2017), on the other hand, conducted several latent profile analyses to identify reading profiles in different age groups and found that, while profiles in elementary grades show heterogenous deficits, profiles in higher age groups mostly showed a high, medium, low pattern of parallel profiles. This implies that the identified patterns are subject to various developmental trajectories and thereby not necessarily stable over longer timeframes.

In addition to the specific profiles identified, we observed that the deviation from the mean for the good and poor readers is especially pronounced for reading comprehension. This is in line with the presumed multiplicative relation between decoding and language comprehension in the simple view of reading (Hoover & Gough, 1990). Thus, when the two skills of decoding and language comprehension are both not yet developed enough or both well developed, they have an even stronger effect on the resulting reading comprehension level than when only one of the two prerequisite skills is high or low.

While this was an exploratory study, the class-specific regression weights of specific instructional foci still provide some implications for reading instruction:

- (1) The negative effect of vocabulary training on the reading comprehension of poor comprehenders might suggest that learners who struggle in reading comprehension have trouble expanding their vocabulary by just reading and inferring meaning from context (Duff et al., 2015; Suk, 2017). However, given the exploratory nature of the analyses, replications of these results are clearly needed to support this interpretation. This again highlights the strength of the multivariate approach. Simply looking at the specific deficits of children classified as poor comprehenders would make vocabulary and syntactic training the straightforward choice (Galuschka et al., 2014). By instead teaching comprehension skills to complement their already strong decoding abilities, teachers can enable those children to improve their vocabulary on their own, while also increasing their reading comprehension level (Share, 1999 ; Verhoeven et al., 2011).
- (2) In contrast, for children who already have strong decoding and comprehension abilities, the most valuable focus seems to be on vocabulary extension, even though they already possess a good vocabulary. This is plausible since it is possible that the best way to improve a completely automated reading process is to add even more words to the mental lexicon.
- (3) The observation that fostering reading motivation did not have any significant effect on the development of reading comprehension, regardless of profile, can potentially be explained by mediating mechanisms. If the positive effects of reading motivation on reading comprehension are, for example, mediated by an increased frequency of reading outside the school context, they may take longer to manifest than the time frame of this study was able to capture (Guthrie et al., 1999; Retelsdorf et al., 2011).

It is important to note that, at a more general level, our results indicate a need for individualized or differentiated reading instruction. A uniform instructional focus for an entire class of students is certain to be a wasted opportunity for some of them. The strong differential effectiveness of specific instructional approaches depending on measurable multivariate aptitude profiles implies a need for stronger individualization or at least differentiation of instruction—this is in line with previous research on individualized reading instruction (Connor et al., 2007, 2009). Basing instructional adaptations on multivariate aptitude profiles, rather than specific univariate deficits, potentially enables even more effective individualization that also builds on students' individual strengths.

This is especially apparent when looking at vocabulary as a learning prerequisite. The groups of “good readers” and “poor decoders” have almost identically strong vocabularies, but—for the “good readers”—putting a focus on improving it further seems to be the most effective use of instructional time, while for the “poor comprehenders,” a focus on advanced reading abilities such as text comprehension seems to be more prudent. By just looking at childrens' vocabulary measures and their interaction with the different instructional foci, both of these effects would have been

missed (see Table 2). A similar discrepancy can be found when looking at the results of a focus on vocabulary training: The variable-centered analyses paint a one-sided picture of its interaction with different learning prerequisites—implying that no learners specifically benefit from it—while the person-centered approach reveals a subgroup of students, i.e., the “good readers,” who greatly benefit from it.

Limitations

The main limitation of the present study is its primarily exploratory nature. This implies that both the identified profiles and the observed interaction with instructional foci need to be replicated before they can be used to inform specific instructional approaches. Regarding the profile analysis, it should be mentioned that the selected indicator variables are not completely exhaustive indicators of language comprehension. However, both vocabulary knowledge (Tunmer & Chapman, 2012) and syntactic comprehension (Tilstra et al., 2009) have been identified as important aspects of language comprehension.

The teacher self-reports might be biased by educationally desirable response tendencies, leading to an overestimation of the amount of focus they put on specific instructional practices (e.g., Tetzlaff et al., 2022). This is indicated by consistently positive correlations ($r = .30$ to $.55$) between the different instructional foci. If teachers managed to report their instructional practices in a reliable manner, it would result in lower intercorrelations between the amount of focus they put on each of the individual instructional practices. For example, if teachers who teach more advanced reading during a given period tend to teach less vocabulary because there is only enough time for one of them, then this should result in a negligible or even negative correlation between these two instructional foci. The fact that we found positive correlations between all instructional foci indicates that on average, teachers might tend to engage in consistently positively biased response behavior across all instructional foci. Prior research indicates that, although teacher self-reports about teaching practices are related to actual classroom observations (Mayer, 1999), such self-reports might still be biased by social desirability aspects (Wubbels et al., 1992). However, for our focal interpretations, this might not pose a significant issue because we applied a multiple regression approach in our analysis of differential effectiveness. This approach should correct for bias by controlling for the shared positive covariance among the predictor variables that indicate the different instructional foci.

In addition to response biases, it is conceivable that the teachers differ in their understanding of what is meant by terms like *advanced reading abilities*. However, in our assessments, we labeled the instructional foci with simple descriptions that were in accordance with the teachers' official instructional guidelines and curricula. Consequently, it should be rather unlikely that the positive correlations of the instructional foci are largely caused by linguistic ambiguities. To further explore and control for potential response biases, future studies with larger samples of teachers could specify a multilevel mixture in the model (see Flunger et al., 2021; Vermunt, 2008). This would allow for better differentiation of the general effects of a specific focus versus teachers' specific implementations.

Another limitation of the present study is the relatively small sample size, especially for the subgroup of “poor readers.” Future research could aim at sampling larger groups of “poor readers” in order to derive meaningful predictions for this group—which is arguably the group most in need of adaptive instruction. This can be done either by increasing the sample size in general or by selecting a population with a higher likely proportion of poor readers.

Future Directions

Building up on our findings, an important next step would be follow-up studies that adopt a more confirmatory approach with clear testable a priori hypotheses. In the best case, the analytic approach could be taken up by a randomized control study in which teachers are randomly assigned to place their focus on a specific aspect of reading instruction (randomly assigning students to teachers would probably not be realistic) or students get placed in specific programs, focusing on one of the instructional aspects. This would not only resolve potential ambiguities in the instructional focus measure but also eliminate any hidden confounders at the teacher level. Even if an experimental manipulation is not possible, using different ways to assess the instructional foci, such as classroom observations, analyses of teaching materials or teacher interviews or even a combination of these approaches, would strengthen the validity of the interpretation of these critical indicators.

Another important next step would be to test the more general applicability of our person-centered approach in areas within and beyond reading. Within reading, future research could for example examine the connection between different instructional foci and the development of language comprehension and decoding ability. This would also provide additional information on whether an improvement in the component skill mediates the effect of instruction on reading comprehension gains. Relatedly, it might be informative to broaden the range of learner characteristics that make up the multivariate learner model, including affective/motivational dispositions as well as personality traits (Ackerman, 2003). Examples of multivariate constructs that might lend themselves to building aptitudes and modeling differential effectiveness in combination with the present approach include executive functions (Miyake et al., 2000), self-regulation (Grunschel et al., 2013), and affective/motivational variables such as different kinds of goal orientations (Wolters, 2004).

One could argue that—from a practical point of view—the least amount of indicators that allow the profiles to inform differential effectiveness of treatment parameters would be preferable, as this would—in theory—not only reduce the amount of data to be assessed but also reduce the amount of different instructional approaches that need to be employed. However, from a theoretical point of view, it might be worthwhile to include additional parameters as long as they increase the quality of profile estimation (Wurpts & Geiser, 2014), as this additional information might uncover further mechanisms by which the instructional parameters interact with the individual learners.

In addition to broadening the scope of aptitudes in these regards, similar approaches could be taken for outcome variables. The recent study by Grimm et al. (2023) demonstrates that, in the investigation of differential effectiveness, latent profile analysis also lends itself well to the modeling of multivariate learning outcomes. Another promising extension of the current approach might be the repeated assessment of indicator variables. This would allow for a more dynamic conceptualization of aptitude profiles and their interaction with specific kinds of instruction. Reinhold et al. (2020) demonstrated, for example, how process data can be used to build profiles of students with different patterns of engagement. Such an approach would also enable an investigation into whether and how teachers adapt their instruction year to changing learner prerequisites over the school year. This could be further extended by relaying information about the multivariate aptitude profiles of their students to teachers (either in a dynamic way via formative assessment procedures or based on single measurement points) and observing if teachers adapt their instruction based on that information and how this affects learning. Over one school year, the assessed multivariate learner prerequisites may change in interaction with the learning process. We only assessed them once at the beginning of the school year, leading to a potential mismatch when learners make rapid gains in one of these areas in the first few weeks or months of instruction. A more dynamic measurement approach would allow for better differentiation of these effects, as well as a better understanding of the temporal dynamics behind them (Tetzlaff et al., 2021).

Conclusions

In this study, we were able to show that profiles of multivariate aptitudes can be used to explain the differential effectiveness of treatments above and beyond univariate conceptualizations, at least in the domain of reading. The person-centered approach circumvents the exorbitant power requirements and interpretational complexity involved in analyzing higher-order interactions in variable-centered multiple regression models. The differential effectiveness of instructional parameters that do not show a significant main effect across all learners suggests that those parameters need to be selectively adapted to specific learners. Our analytic approach appears promising for identifying differential effectiveness, potentially providing a way to overcome the long-standing methodological bottleneck in this area across a variety of educational domains.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10648-023-09830-5>.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability All data and analysis code have been made publicly available and can be accessed at https://osf.io/a97gv/?view_only=73f3249ac61f4b618415d3116e1b164f.

Declarations

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ackerman, P. L. (2003). Aptitude complexes and trait complexes. *Educational Psychologist*, 38(2), 85–93. https://doi.org/10.1207/S15326985EP3802_3
- Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus Web Notes*, 21(2), 1–22.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bauer, D. J., & Shanahan, M. J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. In T. Little, J. Bovaird, & N. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 255–283). Routledge.
- Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose (s)? *Educational Psychology Review*, 33(4), 1675–1715.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16. <https://doi.org/10.3102/0013189X013006004>
- Bracht, G. H. (1970). Experimental factors related to aptitude-treatment interactions. *Review of Educational Research*, 40(5), 627–645. <https://doi.org/10.2307/1169460>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5–51.
- Clunies-Ross, P., Little, E., & Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behaviour. *Educational Psychology*, 28(6), 693–710.
- Connor, C. M. D., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science*, 315(5811), 464–465. <https://doi.org/10.1126/science.1134513>
- Connor, C. M. D., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., Underwood, P., & Morrison, F. (2009). Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. *Child Development*, 80(1), 77–100.
- Constas, M. A., & Sternberg, R. J. (2013). *Translating theory and research into educational practice: Developments in content domains, large-scale reform, and intellectual capacity*. Routledge. <https://doi.org/10.4324/9780203726556>
- Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, 43(3), 161–173. <https://doi.org/10.1080/00461520802178466>
- Coyne, M. D., McCoach, D. B., Ware, S., Austin, C. R., Loftus-Rattan, S. M., & Baker, D. L. (2019). Racing against the vocabulary gap: Matthew effects in early vocabulary instruction and intervention. *Exceptional Children*, 85(2), 163–179. <https://doi.org/10.1177/0014402918789162>

- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671.
- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*(2), 116–127. <https://doi.org/10.1037/h0076829>
- Cronbach, L. J., & Snow, R. E. (1981). *Aptitudes and instructional methods: A handbook for research on interactions*. Ardent Media.
- Decristan, J., Fauth, B., Kunter, M., Büttner, G., & Klieme, E. (2017). The interplay between class heterogeneity and teaching quality in primary school. *International Journal of Educational Research*, *86*, 109–121.
- Deno, S. L. (1990). Individual differences and individual difference. *The Journal of Special Education*, *24*(2), 160–173. <https://doi.org/10.1177/002246699002400205>
- Dockterman, D. (2018). Insights from 200+ years of personalized learning. *Npj Science of Learning*, *3*(1), 1–6. <https://doi.org/10.1038/s41539-018-0033-x>
- Duff, D., Bruce Tomblin, J., & Catts, H. (2015). The influence of reading on vocabulary growth: A case for a Matthew effect. *Journal of Speech, Language, and Hearing Research*, *58*(3), 853–864. https://doi.org/10.1044/2015_JSLHR-L-13-0310
- Edelsbrunner, P., Flaig, M., & Schneider, M. (2023). A simulation study on latent transition analysis for examining profiles and trajectories in education: Recommendations for fit statistics. *Journal of Research on Educational Effectiveness*, *16*(2), 350–375.
- Ehm, J. H., Schmitterer, A. M., Nagler, T., & Lervåg, A. (2023). The underlying components of growth in decoding and reading comprehension: Findings from a 5-year longitudinal study of German-speaking children. *Scientific Studies of Reading*, 1–23.
- Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, *2*(1), 1–44.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für pädagogische Psychologie*.
- Ferguson, S. L., Moore, G., & Moore, E. W. G. (2020). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development*, *44*(5), 458–468.
- Fleiss, J. L. (1986). Reliability of measurement. In *The design and analysis of clinical experiments* (pp. 1–32). Wiley.
- Flunger, B., Trautwein, U., Nagengast, B., Lüdtke, O., Niggli, A., & Schnyder, I. (2021). Using multilevel mixture models in educational research: An illustration with homework research. *The Journal of Experimental Education*, *89*(1), 209–236.
- Foorman, B. R., Petscher, Y., Stanley, C., & Trueman, A. (2017). Latent profiles of reading and language and their association with standardized reading outcomes in kindergarten through tenth grade. *Journal of Research on Educational Effectiveness*, *10*(3), 619–645.
- Freebody, P., & Anderson, R. C. (1983). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Literacy Research*, *15*(3), 19–39. <https://doi.org/10.1080/10862968309547487>
- Freedman, D. A. (2006). On the so-called “Huber sandwich estimator” and “robust standard errors”. *The American Statistician*, *60*(4), 299–302.
- Fuchs, D., Kearns, D. M., Fuchs, L. S., Elleman, A. M., Gilbert, J. K., Patton, S., Peng, P., & Compton, D. L. (2019). Using moderator analysis to identify the first-grade children who benefit more and less from a reading comprehension program: A step toward aptitude-by-treatment interaction. *Exceptional Children*, *85*(2), 229–247. <https://doi.org/10.1177/0014402918802801>
- Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, J., Malone, A., Hamlett, C. L., Jordan, N. C., Gersten, R., Siegler, R. S., & Chngas, P. (2014). Does working memory moderate the effects of fraction intervention? An aptitude-treatment interaction. *Journal of Educational Psychology*. <https://doi.org/10.1037/a0034341>
- Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS One*, *9*(2), e89900. <https://doi.org/10.1371/JOURNAL.PONE.0089900>
- Glück, C. W. (2011). *Wortschatz- und Wortfindungstest für sechs- bis zehnjährige (WWT 6-10)* [Vocabulary and word finding test for six to ten year olds (WWT 6-10)]. Elsevier.

- Grimm, H., Edelsbrunner, P. A., & Möller, K. (2023). Accommodating heterogeneity: The interaction of instructional scaffolding with student preconditions in the learning of hypothesis-based reasoning. *Instructional Science*, 51(1), 103–133.
- Grunschel, C., Patrzek, J., & Fries, S. (2013). Exploring different types of academic delayers: A latent profile analysis. *Learning and Individual Differences*, 23, 225–233.
- Gudicha, D. W., & Vermunt, J. K. (2013). Mixture model clustering with covariates using adjusted three-step approaches. In B. Lausen, D. Van den Poel, & A. Utzsch (Eds.), *Algorithms from and for nature and life: Classification and data analysis* (pp. 87–94). Springer, Cham. https://doi.org/10.1007/978-3-319-00035-0_8
- Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading*, 3(3), 231–256. https://doi.org/10.1207/s1532799xssr0303_3
- Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation : An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 621–638. <https://doi.org/10.1080/10705511.2017.1402334>
- Harring, J. R., & Hodis, F. A. (2016). Mixture modeling: Applications in educational psychology. *Educational Psychologist*, 51(3-4), 354–367.
- Hessian Ministry of Education (2021). *Rahmenplan Grundschule*. <https://grundschule.bildung.hessen.de/rahmenplan/Rahmenplan.pdf>
- Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences*, 66, 4–15. <https://doi.org/10.1016/j.lindif.2017.11.001>
- Hjetland, H. N., Brinchmann, E. I., Scherer, R., Hulme, C., & Melby-Lervåg, M. (2020). Preschool pathways to reading comprehension: A systematic meta-analytic review. *Educational Research Review*, 30, 100323.
- Hjetland, H. N., Lervåg, A., Lyster, S. A. H., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (2019). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology*, 111(5), 751–763. <https://doi.org/10.1037/edu0000321>
- Hofer, S. I., Schumacher, R., Rubin, H., & Stern, E. (2018). Enhancing physics learning with cognitively activating instruction: A quasi-experimental classroom intervention study. *Journal of Educational Psychology*, 110(8), 1175–1191. <https://doi.org/10.1037/EDU0000266>
- Hooper, S. R., Wakely, M. B., de Kruif, R. E., & Swartz, C. W. (2006). Aptitude–treatment interactions revisited: Effect of metacognitive intervention on subtypes of written expression in elementary school students. *Developmental Neuropsychology*, 29(1), 217–241.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127–160. <https://doi.org/10.1007/BF00401799>
- Hoover, W. A., & Tunmer, W. E. (2018). The simple view of reading: Three assessments of its adequacy. *Remedial and Special Education*, 39(5), 304–312. <https://doi.org/10.1177/0741932518773154>
- Hoover, W. A., & Tunmer, W. E. (2022). The primacy of science in communicating advances in the science of reading. *Reading Research Quarterly*, 57(2), 399–408. <https://doi.org/10.1002/rrq.446>
- Hunt, D. E. (1975). Person-environment interaction: A challenge found wanting before it was tried. *Review of Educational Research*, 45(2), 209–230. <https://doi.org/10.2307/1170054>
- Jiang, D., Kalyuga, S., & Sweller, J. (2018). The curious case of improving foreign language listening skills by reading rather than listening: An expertise reversal effect. *Educational Psychology Review*, 30(3), 1139–1165. <https://doi.org/10.1007/s10648-017-9427-1>
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19(4), 509–539. <https://doi.org/10.1007/s10648-007-9054-3>
- Kendeou, P., McMaster, K. L., & Christ, T. J. (2016). Reading comprehension: Core components and processes. *Policy Insights From the Behavioral and Brain Sciences*, 3(1), 62–69.
- Kopcha, T. J., & Sullivan, H. (2007). Self-presentation bias in surveys of teachers' educational technology practices. *Educational Technology Research and Development*, 55(6), 627–646. <https://doi.org/10.1007/s11423-006-9011-8>
- Kuşdemir, Y., & Bulut, P. (2018). The relationship between elementary school students' reading comprehension and reading motivation. *Journal of Education and Training Studies*, 6(12), 97–110. <https://doi.org/10.11114/jets.v6i12.3595>

- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11(4), 815–852.
- Lenhard, W., & Schneider, W. (2006). *ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler*. Hogrefe.
- Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, 89(5), 1821–1838.
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767–778.
- Lonigan, C. J., Burgess, S. R., & Schatschneider, C. (2018a). Examining the simple view of reading with elementary school children: Still simple after all these years. *Remedial and Special Education*, 39(5), 260–273.
- Lonigan, C. J., Goodrich, J. M., & Farver, J. M. (2018b). Identifying differences in early literacy skills across subgroups of language-minority children: A latent profile analysis. *Developmental Psychology*, 54(4), 631.
- Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person-and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(2), 191–225.
- Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis*, 21(1), 29–45. <https://doi.org/10.3102/01623737021001029>
- McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, 22(1), 114.
- Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research*, 53(2), 253–279.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100.
- Moll, K., & Landerl, K. (2010). *Lese- und Rechtschreibtest (SLRT-II)*. Huber.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535–569.
- Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A latent transition mixture model using the three-step specification. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 439–454.
- Nylund-Gibson, K., Grimm, R. P., & Masyn, K. E. (2019). Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(6), 967–985.
- Preacher, K. J., & Sterba, S. K. (2019). Aptitude-by-treatment interactions in research on educational interventions. *Exceptional Children*, 85(2), 248–264.
- R Core Team. (2021). *R: A language and environment for statistical computing (4.0.2)*. R Foundation for Statistical Computing <https://www.r-project.org/>
- Reinhold, F., Strohmaier, A., Hoch, S., Reiss, K., Böheim, R., & Seidel, T. (2020). Process data from electronic textbooks indicate students' classroom engagement. *Learning and Individual Differences*, 83, 101934.
- Retelsdorf, J., Köller, O., & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction*, 21(4), 550–559. <https://doi.org/10.1016/J.LEARNINSTRUC.2010.11.001>
- Schiefele, U., Stutz, F., & Schaffner, E. (2016). Longitudinal relations between reading motivation and reading comprehension in the early elementary grades. *Learning and Individual Differences*, 51, 49–58. <https://doi.org/10.1016/J.LINDIF.2016.08.031>
- Segers, E., & Verhoeven, L. (2003). Effects of vocabulary training by computer in kindergarten. *Journal of Computer Assisted Learning*, 19(4), 557–566. <https://doi.org/10.1046/J.0266-4909.2003.00058.X>
- Seufert, T., Schütze, M., & Brünken, R. (2009). Memory characteristics and modality in multimedia learning: An aptitude-treatment-interaction study. *Learning and Instruction*, 19(1), 28–42. <https://doi.org/10.1016/j.learninstruc.2008.01.002>
- 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>

- Siegmüller, J., Kauschke, V. M. S., & Bittner, D. (2011). *Test zum Satzverstehen von Kindern*. Elsevier, Urban & Fischer Verlag.
- Snow, R. E. (1987). Aptitude complexes. In R. Snow & M. Farr (Eds.), *Aptitude, learning, and instruction* (pp. 11–34). Routledge.
- Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A., & Clausen, J. A. (1950). *Measurement and prediction*. Princeton University Press.
- Subban, P. (2006). Differentiated instruction: a research basis. *International Education Journal*, 7(7), 935–947.
- Suk, N. (2017). The effects of extensive reading on reading comprehension, reading rate, and vocabulary acquisition. *Reading Research Quarterly*, 52(1), 73–89. <https://doi.org/10.1002/RRQ.152>
- Tambyraja, S. R., Schmitt, M. B., Farquharson, K., & Justice, L. M. (2015). Stability of language and literacy profiles of children with language impairment in the public schools. *Journal of Speech, Language, and Hearing Research*, 58(4), 1167–1181. https://doi.org/10.1044/2015_JSLHR-L-14-0197
- Tetzlaff, L., Hartmann, U., Dumont, H., & Brod, G. (2022). Assessing individualized instruction in the classroom: Comparing teacher, student, and observer perspectives. *Learning and Instruction*, 82, 101655.
- Tetzlaff, L., Schmiedek, F., & Brod, G. (2021). Developing personalized education: A dynamic framework. *Educational Psychology Review*, 33(3), 863–882. <https://doi.org/10.1007/S10648-020-09570>
- Tilstra, J., McMaster, K., Van Den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading*, 32(4), 383–401. <https://doi.org/10.1111/J.1467-9817.2009.01401.X>
- Torppa, M., Tolvanen, A., Poikkeus, A. M., Eklund, K., Lerkkanen, M. K., Leskinen, E., & Lyytinen, H. (2007). Reading development subtypes and their early characteristics. *Annals of Dyslexia*, 57(1), 3–32.
- Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, 45(5), 453–466. <https://doi.org/10.1177/00222194111432685>
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- Vaughn, S., Roberts, G., Capin, P., Miciak, J., Cho, E., & Fletcher, J. M. (2019). How initial word reading and language skills affect reading comprehension outcomes for students with reading difficulties. *Exceptional Children*, 85(2), 180–196. <https://doi.org/10.1177/0014402918782618>
- Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, 15(1), 8–25.
- Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, 17(1), 33–51.
- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4), 450–469.
- Wolff, U. (2010). Subgrouping of readers based on performance measures: A latent profile analysis. *Reading and Writing*, 23(2), 209–238. <https://doi.org/10.1007/S11145-008-9160-8>
- Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology*, 96(2), 236.
- Wubbels, T., Brekelmans, M., & Hoymayers, H. P. (1992). Do teacher ideals distort the self-reports of their interpersonal behavior? *Teaching and Teacher Education*, 8(1), 47–58. [https://doi.org/10.1016/0742-051X\(92\)90039-6](https://doi.org/10.1016/0742-051X(92)90039-6)
- Wurpts, I. C., & Geiser, C. (2014). Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. *Frontiers in Psychology*, 5, 920.
- Ziegler, E., Edelsbrunner, P. A., & Stern, E. (2020). The benefit of combining teacher-direction with contrasted presentation of algebra principles. *European Journal of Psychology of Education*. <https://doi.org/10.1007/s10212-020-00468-3>