Effectiveness of Professional Development for Language-Supportive Teaching: Insights From Employing Computational Linguistic Analysis Methods

Birgit Heppt, Denise Löfflad, Sofie Henschel, Katrin Gabler, Ilonca Hardy & Detmar Meurers

Zusammenfassung

Die vorliegende Studie untersucht das Potenzial computerlinguistischer Analyseverfahren zur Evaluation der Effekte von Lehrkräftefortbildungen. Entsprechende Methoden ermöglichen die automatisierte Analyse einer Vielzahl sprachlicher Merkmale und dürften daher eine vertiefte Analyse der mündlichen Unterrichtssprache von Lehrkräften erleichtern. Auf Grundlage eines positiv evaluierten Professionalisierungsansatzes (z. B. Heppt et al., 2022) wurde untersucht, ob Lehrkräfte, die für die Umsetzung fachintegrierter Sprachbildung im Sachunterricht der Grundschule fortgebildet wurden (Interventionsgruppe [IG], $n = 9$), ihren Schüler*innen einen variableren und anspruchsvolleren Sprachinput anboten als Lehrkräfte der Kontrollgruppe (KG; $n = 19$). Anhand von Transkripten von zwei videografierten Sachunterrichtsstunden je Lehrkraft (Thema 1: Schwimmen und Sinken, Thema 2: Verdunstung und Kondensation) fanden wir nur wenige Merkmale, die für eine elaboriertere und komplexere Sprachverwendung der IG im Vergleich zur KG sprechen. Die Gruppenunterschiede waren überdies bei Thema 2 weniger stark ausgeprägt als bei Thema 1, was darauf hindeutet, dass sich die Fortbildungsteilnahme nur geringfügig auf den mündlichen Sprachgebrauch der Lehrkräfte während des Sachunterrichts auswirkte.

Schlagworte: Lehrkräftefortbildung, Sprachbildung, Grundschule, Sachunterricht, sprachliche Komplexität, computerlinguistische Analyseverfahren

Abstract

The present study aims at exploring the potential of using computational linguistic analysis methods for examining professional development (PD) effects. As respective methods allow for an automatic analysis of a wide range of linguistic features, they may facilitate an in-depth analysis of teachers' oral language use in classroom discourse. Building on a positively evaluated PD approach (e.g., Heppt et al., 2022), we investigated whether in-service teachers who were trained for language-supportive teaching in elementary school science classes (intervention group [IG], $n = 9$) provide their students with a more varied and more elaborate language input than teachers from the control group (CG; $n = 19$). Based on transcripts of video recordings of two science lessons per teacher (Topic 1: floating and sinking, Topic 2: evaporation and condensation), we found only a small number of features that point to more sophisticated and complex language use in the IG than in the CG. Moreover, group differences were less pronounced for Topic 2 than for Topic 1, suggesting that PD participation had only small effects on teachers' oral language use during science instruction.

Keywords: Teacher professional development, language support, elementary school, science education, linguistic complexity, computational linguistic analysis methods

1. Introduction

Language-supportive teaching that fosters both domain-specific learning and language development is increasingly seen as a general teaching principle across subjects (e.g., Becker-Mrotzek & Woerfel, 2020; Prediger & Hardy, 2023). Although considered beneficial for all students, such classroom instruction is particularly aimed at students who are at risk of falling behind. Among them are students with an immigrant background, often growing up as dual language learners, who have repeatedly been shown to lag behind their peers without immigrant background in academic achievement and (academic) language proficiency (Henschel et al., 2022; Ludewig et al., 2022). Similarly, students from families with low socioeconomic status (SES) consistently perform, on average, below their peers from high-SES families (e.g., Sachse et al., 2022). Both student groups are likely to have only limited opportunities for acquiring the register of schooling within their families. Children's language development depends on the amount and quality of language input at home, which, in turn, varies with family SES. Children from high-SES families, thus, typically outperform their counterparts from low-SES families in their mastery of academic language, helping them to benefit from classroom discourse, accomplish written assignments, or understand specialized texts (e.g., Volodina et al., 2021). Students with an immigrant background face similar obstacles as they often grow up in families with low SES (e.g., Henschel et al., 2022). Moreover, many of them have limited access to the language of instruction within their homes, thus further hampering their mastery of the language of schooling (for an overview, see Heppt & Schröter, 2023). With the aim of adapting to these students' needs, language-supportive teaching is considered a potential remedy for tackling educational inequalities.

The need for integrating language support into regular classroom teaching is widely accepted, resulting in an expansion of course offerings on language support and second language acquisition in university teacher training in Germany (Paetsch & Heppt, 2021). While an increasing number of German elementary school teachers participates in such courses during university teacher training, many of them are still ill-prepared for providing subject-integrated language support in their daily classroom instruction (Henschel & Heppt, 2024). This highlights the importance of effective professional development (PD).

In investigating the effectiveness of PD for language-supportive teaching, research frequently focuses on teachers' classroom practice (e.g., Gabler et al., 2024; Heppt et al., 2022; van Dijk et al., 2019). Based on laborious coding of selected language-support strategies (e.g., use of language-supportive questions) or highly inferential ratings of the overall quality of language support, studies have pointed to the general effectiveness of PD for developing teachers' classroom practice (for a meta-analysis, see Kalinowski et al., 2020).

With the present study, we aim at exploring the potential of using computational linguistic analysis methods for examining PD effects. As respective methods allow for an automatic analysis of a wide range of linguistic features, they may facilitate an in-depth analysis of teachers' oral language use in classroom discourse (cf. Weiss et al., 2022). Building on a PD approach that has been positively evaluated regarding participants' knowledge on language support (Heppt et al., 2022) and their use of specific language-support strategies in science instruction (e.g., the use of language-supportive questions; Gabler et al., 2024), this study focuses on teachers' oral language input. Specifically, we investigate whether, upon completing the PD, German elementary school teachers use more stimulating and sophisticated oral language in their science instruction than teachers who did not take part in the PD.

2. Theoretical and Empirical Background

2.1. Language-Supportive Classroom Instruction

Effective language-supportive classroom instruction helps students attain domain-specific learning goals, while developing the necessary academic language skills. In elementary school science classes, for instance, experiments on the floating and sinking of objects are clearly aimed at developing students' conceptual knowledge on phenomena like water displacement or water pressure. Yet, in order to construct this type of knowledge in a co-constructive process, students need to be able to formulate and justify assumptions and to describe and explain their observations (e.g., Vorholzer & Aufschnaiter, 2019). These language functions form a core part of the academic language register (e.g., Bailey et al., 2007; Prediger & Hardy, 2023). Performing language functions such as "hypothesizing" or "justifying" requires the precise understanding of their respective meaning. In addition, students need knowledge of adequate linguistic structures (e.g., the correct use of causal connectives such as "therefore" or "due to") and domain-specific academic vocabulary (e.g., "to displace", "wax", "Styrofoam"). As an important prerequisite for learning, language functions need to be systematically developed in classroom instruction, along with the underlying lexical and syntactical skills.

This can best be achieved by cognitively activating instruction that engages students in higher-order thinking and connects concepts with activities (e.g., by conducting experiments), as such learning environments typically offer multiple opportunities for using language in meaningful contexts (e.g., Bravo & Cervetti, 2014). In using language functions and expanding their vocabulary knowledge, students should be assisted through language-support strategies (e.g., open-ended questions, rich and elaborate language input, linguistic feedback that adequately expands the students' utterances; e.g., Gabler et al., 2020; Heppt et al., 2022; Mahan, 2020). The linguistic scaffolding approach (Gibbons, 2002), which builds on the theory of social learning (cf. Wood et al., 1976), considers these linguistic aids as a scaffold that helps students master linguistically demanding tasks. These aids are adapted to students' linguistic needs and gradually reduced as the students increasingly gain proficiency in academic language. The linguistic scaffolding approach has been proven effective in promoting students' domain-specific knowledge with no pronounced differences across student groups (e.g., multilingual and monolingual learners; Prediger & Neugebauer, 2021; Prediger & Wessel, 2017). Research on its effectiveness for (academic) language development has mostly been conducted in the United States (US) with a focus on students with limited language proficiency (English Language Learners; ELLs). Overall, results indicate that integrating inquiry-based science instruction with the adaptive use of a range of language-support strategies benefits ELLs' academic language proficiency (e.g., Bravo & Cervetti, 2014; Llosa et al., 2016).

2.2. Language Input and its Associations with Content- and Language Learning

The quality and amount of oral language input are among the fundamental drivers of children's language development. Since the seminal study by Hart and Risely (1995) on the huge socioeconomic differences in the number of child-directed words, amounting to a "30-million-word gap" for children from high vs. low-SES families by the age of four, numerous studies have highlighted the relation between caregivers' language input and their child's language proficiency. In line with these findings, the amount and quality of a (preschool) teacher's oral language use are also considered an essential language-support strategy in instructional settings (e.g., Gabler et al., 2020; Kane et al., 2023).

With the aim of acting as language role models, this language-support strategy requires (preschool) teachers to provide frequent, rich, and elaborate language input in instruction. This can be achieved by using thinking-aloud techniques or by mapping one's own or students' actions with language (e.g., "I have a wooden dice in my left hand and a metal dice in my right hand. Now, I put the wooden dice into the water basin."). In doing so, teachers should try and use important general and domain-specific vocabulary, ideally in multiple contexts and by contrasting them with other words (e.g., "Did we hypothesize this or did we establish this? We established this in the end. So this is for sure."), contributing to students' enhanced and differentiated vocabulary knowledge.

Prior research on teacher's oral language use has been conducted in preschool settings and, to a lesser extent, in elementary school. Overall, this research showed that preschool teachers' use of elaborate language can increase students' domain-specific learning and language development (e.g., Kane et al., 2023; Studhalter et al., 2021). However, (preschool) teachers tend to use this high-quality language input rather infrequently (see however, Weiss et al., 2022). Studhalter et al. (2021), for instance, found that preschool teachers' use of domain-specific vocabulary (e.g., "wax", "clay", "iron") during a 4-week learning unit on the topic "floating and sinking" significantly predicted children's conceptual learning gains. Focusing on the interplay between the quality of preschool teachers' talk and students' language development, Dickinson and Porche (2011) conducted a longitudinal study from preschool to fourth grade. Among the indicators used for evaluating the quality of preschool teachers' oral language were sophisticated vocabulary (i.e., low-frequency words) and utterances aimed at focusing children's attention, correcting or expanding their oral expressions. The authors found that each of these indicators of preschool teachers' language quality contributed to students' reading skills in Grade 4 (Dickinson & Porche, 2011).

Despite its pivotal role for student learning, children seem to receive relatively low amounts of high-quality language input during regular classroom teaching in elementary school. Based on videotaped classroom observations of five upper elementary classrooms in the US, Ernst-Slavit and Mason (2011) found that less than 12% of teachers' oral language input across classrooms and subjects can be classified as "academic language", as reflected in vocabulary (e.g., general and domain-specific vocabulary), grammar (e.g., syntactically long and complex sentences with clause connectives) and discourse (e.g., factual and information-dense style). In a similar vein, a Dutch study aimed at assessing elementary school teachers' use of academic language in whole-classroom discourse in mathematics classrooms in Grades 1 and 2 (Dokter et al., 2017). Drawing on transcripts of two eight-minute sequences of classroom instruction per teacher, the authors found that all teachers used math-specific language to some extent. However, the overall lexical and grammatical complexity of teacher talk was rather low. In sum, prior research suggests that teachers do not deliberately use their oral language input in classroom discourse as a means for modeling students' (academic) language development, pointing to the need for effective teacher PD.

2.3. Effectiveness of Teacher PD for Language-Supportive Teaching

Research on teacher PD in general (e.g., Darling-Hammond et al., 2009; Lipowsky & Rzejak, 2015) as well as the emerging literature on PD for language-supportive teaching (for an overview, see Kalinowski et al., 2020) have identified key characteristics of effective teacher PD. Importantly, teacher PD that helps teachers gain knowledge and skills for integrating language-support strategies into regular classroom teaching combines phases of input with opportunities for actively using the newly acquired knowledge (e.g., in role plays or in classroom instruction). During implementation phases, teachers should receive feedback and have the chance to reflect upon their experiences, for instance by analyzing and discussing video-recordings of their own classroom teaching (Piwowar et al., 2013; van Dijk et al., 2019).

PD programs that consider these core principles have, indeed, been proven effective in developing teachers' knowledge and skills for subject-integrated language support (e.g., Babinski et al., 2018; van Dijk et al., 2019). A meta-analysis incorporating ten studies, showed small, albeit statistically non-significant gains in teachers' self-efficacy and self-assessed knowledge regarding language-supportive teaching, whereas larger PD effects occurred for teachers' language-supportive classroom behavior (Kalinowski et al., 2020). Most of the studies included in the meta-analysis were conducted in the US with a focus on ELLs and only few studies used (quasi-)experimental designs with intervention group (IG) and control group (CG).

A couple of more recent (quasi-)experimental studies from Germany and the Netherlands aimed at teachers in mainstream classrooms, helping them to provide language-supportive science instruction for *all* students (Gabler et al., 2020; Henrichs & Leseman, 2014; van Dijk et al., 2019). In one such PD program based on the linguistic scaffolding approach (Gabler et al., 2020), teachers were familiarized with and actively used core language-support strategies in group work and classroom teaching. The PD covered the following language-support strategies: (1) language modeling by providing elaborate and targeted language input, (2) asking language-stimulating questions, (3) giving language-supportive feedback (e.g., by elaborating on or rephrasing students' answers), and (4) shifting students' attention to important vocabulary and sentence structures (e.g., by using visual aids). This PD program, which also forms the basis of the present study, has been proven effective in advancing teachers' knowledge on subject-integrated language support (Heppt et al., 2022). Moreover, after dealing intensively with a science curriculum on "floating and sinking" and the implementation of possible language scaffolds, trained teachers showed a higher overall quality and amount of language-supportive behavior and provided better learning-related feedback in classroom teaching (Heppt et al., 2022). They also used language-stimulating questions more frequently than their counterparts who did not participate in the PD (Gabler et al., 2024). Regarding the overall quality of instructional support, including language-supportive teaching, no group differences were observed when participants were required to transfer their newly acquired knowledge and skills for language-supportive teaching to a different science curriculum (i.e., on "evaporation and condensation"). Interestingly, though, instructional quality of both IG- and CG-teachers was higher when teaching this second topic, possibly reflecting that it was easier to implement with regard to both conceptual demands and classroom organization. At the same time, it should be noted that the two science topics require the use of different science vocabulary and focus on partly different language functions and structures. Hence, the topics may also differ in their potential for modeling sophisticated vocabulary, syntactic structures, and language functions. However, group differences in terms of the amount and quality of teachers' language input have not yet been investigated. Consequently, it remains an open question whether teachers' oral language use differs across topics.

2.4. The Potential of Using Computational Linguistic Analysis Methods for Investigating Teachers' Oral Language Input

In general, research on PD for language-supportive teaching has rarely considered the amount and linguistic sophistication (e.g., the use of low-frequency vocabulary) of teachers' oral language in classroom discourse as an outcome (for exceptions, see Henrichs & Leseman, 2014; van Dijk et al., 2019). At least in part, this is probably due to the enormous amount of time and effort required to assess the frequency of selected features in oral language. In their study on the effectiveness of a short intervention on academic language use in early science instruction, Henrichs and Leseman (2014), for instance, focused on two lexical features: lexical diversity, as measured in the number of word types (i.e., number of different words), and lexical sophistication, as measured in the number of general and domain-specific academic words. The number of word types was determined by counting the number of different words used across all (videotaped and transcribed) lessons. Assessing the number of general and domain-specific vocabulary required the authors to (1) create lists of all words used in the teacher-student conversations, (2) manually double-code all words that were classified as cross-disciplinary vocabulary (e.g., "experiment", "describe") or domain-specific vocabulary (e.g., "air pressure", "force"), and (3) determine interrater-reliability and discuss divergent ratings (Henrichs & Leseman, 2014).

Using computational linguistic analysis methods, by contrast, allows for assessing a wider range of linguistic features automatically in a timely manner. This enables a very fine-grained analysis of teacher talk, specifically incorporating a variety of features that are focused upon in the respective intervention. Moreover, computational linguistic analysis methods come along with maximum accuracy regarding the detection of specific linguistic features in the texts (compared to human raters) and facilitate the implementation of computationally complex and sophisticated features. Whereas determining the number of word types as an indicator of lexical diversity certainly is utterly time-consuming when conducted manually, for example, computational linguistics enable an easy application of methodologically considerably more advanced measures of lexical diversity, resulting in less biased estimates and, thus, higher data quality (McCarthy & Jarvis, 2010; Weiss, 2023). Computational linguistic analysis methods, therefore, are a promising approach for investigating PD effects on teachers' oral language input in science teaching.

3. Research Questions and Hypotheses

Expanding prior findings on a PD approach for language-supportive teaching in elementary school science classes (Gabler et al., 2024; Heppt et al., 2022), the present study focuses on teacher talk in oral classroom interaction.

Specifically, we investigate the following research questions and hypotheses:

(1) Do IG-teachers differ from CG-teachers in the amount and quality (e.g., lexical diversity, lexical elaboration, morphological complexity) of oral language use in elementary school science classrooms after having participated in a PD program for language-supportive teaching in science classrooms? As IG-teachers were intensively trained for using selected language-support strategies in their science classes, including language modeling, we expected them to outperform the CG-teachers in terms of both quantity and quality of language input.

(2) Do these IG-CG-differences in oral language use persist across learning units on two different elementary school science topics (Topic 1: floating and sinking, Topic 2: evaporation and condensation)? Building on prior findings that point to within-group variability in instructional quality across topics (Heppt et al., 2022) and assuming that language demands differ across topics, we do not expect stable IG-CG-differences in teachers' oral language across topics.

4. Method

4.1. Study Design

The present study uses data from the project ProSach ("Professional development training on content-focused language support in elementary school science instruction"; German: "*Professionalisierungsmaßnahmen zur bedeutungsfokussierten Sprachförderung im Sachunterricht der Grundschule*"[1]), in which a newly developed PD program for language-supportive science teaching was evaluated in a quasi-experimental field trial in two German federal states (Berlin and Hesse). The study was conducted over two school years and consisted of a PD phase in the first school year (2016–2017) and an implementation phase in the second school year (2017–2018). During the PD phase, teachers from IG and (waiting) CG received PD on the elementary school science topics "floating and sinking" and "evaporation and condensation", focusing on the content knowledge and pedagogical content knowledge (PCK) needed for teaching these topics in Grade 3 and 4 and familiarizing participants with the lesson plans and teaching material of the learning units[2]. Both PD courses were led by members of the project team who were experienced researchers (PostDocs) and had a background in elementary school science and didactics. The PD courses were conducted face-to-face and had a duration of 5 hours each. They included a range of didactical elements and methods, such as trainer input, hands-on experiments conducted collaboratively in small groups, and plenum discussions.

IG-teachers additionally participated in a comprehensive PD for subject-integrated language support in elementary school science. This PD program was developed along the lines of well-known characteristics of effective PD (e.g., Darling-Hammond et al., 2009; Kalinowski et al., 2020; see Section 2.3), thus combining phases of input with active trials in classroom and subsequent coaching sessions, which included feedback and joint reflection. The PD program was primarily developed and implemented by a member of the

1 The project was funded by the German Federal Ministry of Education and Research (BMBF) through Grant 01JI1602A, awarded to Humboldt-Universität zu Berlin, Germany (principal investigators: Dr. Sofie Henschel, Prof. Dr. Petra Stanat) and through Grant 01JI1602B awarded to Goethe University Frankfurt, Germany (principal investigator: Prof. Dr. Ilonca Hardy). The authors assume full responsibility of the content of the present publication.

2 A third PD referred to the elementary school science topic "education for sustainable development". Due to high dropout rates during the PD phase and, in particular, during the implementation phase, this topic could not be considered in the analyses and, thus, is not described in this chapter (cf. Heppt et al., 2022).

project team who was an experienced researcher (PostDoc) and had a strong background in linguistics and language support. The PD program comprised two modules (Module 1: Basics of language scaffolding, Module 2: Coaching and video-feedback) which took place over roughly half a school year.

Module 1 consisted of three 4-hour face-to-face workshops, familiarizing participants with the basic components of the language scaffolding approach. This approach consists of an extensive phase of language-supportive lesson planning (macro-scaffolding), followed by the actual language-supportive classroom teaching (micro-scaffolding). Building on the contents of the lesson unit on "floating and sinking", IG-teachers, thus, identified the linguistic challenges of the topic and the learning materials (e.g., syntactical structures and vocabulary needed for formulating and justifying hypotheses, domain-specific vocabulary), discussed methods for assessing their students' language skills, and learned how to define appropriate language-related learning goals (macro-scaffolding). Moreover, they were familiarized with a range of well-established language-support strategies for supporting their students' language development in classroom teaching (micro-scaffolding). Amongst others, they were encouraged to model their students' language skills by frequently using core language functions (i.e., hypothesizing, justifying, comparing) and by introducing and explaining important general (e.g., "to verify", "to assume") and domain-specific vocabulary (e.g., "water cycle", "gaseous"), for example by using it in different contexts, contrasting it with words with a different meaning, or juxtaposing morphological derivations (e.g., "to displace" vs. "displacement"; see Section 2.3 for further information on the language-support strategies). Similarly to the PD courses on elementary school science topics, the workshops for subject-oriented language support incorporated a variety of didactical elements, were application-oriented, and fostered participants' collaboration and reflection.

In the application phase in Module 2, each teacher delivered classroom trials of at least two lessons of the curriculum on "floating and sinking" in their regular elementary school science classrooms. We videotaped one lesson per teacher. The classroom trials were followed by (1) one-on-one feedback with the trainer and (2) a video-feedback session in a small group of participants and with the trainer. These video-feedback sessions had a duration of 2 to 3 hours and aimed at identifying successful examples of language support and at reflecting on possible improvements. To complete the PD phase, IG-teachers participated in a final 3-hour whole-group meeting, summarizing the learnings from the small group sessions (for detailed descriptions of the PD program, see Gabler et al., 2020; Heppt et al., 2022).

During the implementation phase in the second school year, teachers from both groups delivered the learning units on “floating and sinking” and “evaporation and condensation” in their regular Grade 3 or Grade 4 science classrooms. The learning unit on “floating and sinking” comprised six double lessons (90 minutes each), the one on “evaporation and condensation” five double lessons. The second double lesson of both topics was videotaped in each classroom, resulting in two videotaped double lessons per teacher. These videotaped lessons served as a basis for determining the quantity and quality of teachers’ classroom talk. After the implementation phase and, thus, at the beginning of the school year 2018–2019, participants from the (waiting) CG were offered a shortened and optimized version of the PD program for language-supportive teaching.

4.2. Participants

The present analyses are based on data from 27 elementary school teachers from 17 schools who participated in the PD phase and implementation phase of the project. Within each school, all participating teachers were assigned to the same quasi-experimental condition. The IG consisted of 9 teachers from 5 schools (M_{age} = 40.33 years, SD = 4.04, 8 female [88.89%], 1 male [11.11%]) and the CG comprised 19 teachers from 8 schools (M_{age} = 42.17 years, SD = 7.82, 14 female [73.68%], 5 male [26.32%]). There were no meaningful group differences in age ($t(19) = -0.39$, $p = .700$, $d = -0.25$) and gender distribution ($\chi^2(1) = 0.84$, $p = 0.36$, $\varphi < .17$). The groups did not differ in the number of teachers who were trained for teaching in elementary schools ($\chi^2(1) = 2.22$, $p = 0.136$, $\varphi < .29$), but more IG-teachers than CG-teachers had attended courses on German as a second language or language support during university teacher training ($\chi^2(1) = 5.03$, $p = 0.025$, $\varphi < .45$). Moreover, whereas the IG consisted of teachers from Berlin only, the CG was predominantly located in Hesse ($\chi^2(1) = 20.50$, $p < .001$, $\varphi < .86$). This uneven distribution of IG and CG across states resulted from difficulties in the recruiting process and teacher dropout during the PD phase in Berlin. This made it necessary to additionally recruit teachers from Hesse, who, due to time constraints during the ongoing project, could only be assigned to the less time-consuming PD in the CG (for further details on sample attrition and comparability of groups, see Heppt et al., 2022).

Upon PD completion and, thus, before teaching the elementary school science topics to their science classrooms during the implementation phase, IG teachers (as expected) outperformed CG teachers in their knowledge on language-supportive teaching ($t(18) = 0.96$, $p = .174$, $d = 0.64$). Yet, teachers from both groups were on par in their PCK on floating and sinking ($t(20) = -0.06$, $p = .951$, $d = -0.03$) and evaporation and condensation ($t(17) = 0.43$, $p = .672$, $d = 0.21$).

The 27 teachers taught 489 elementary school students from Grades 3 ($n = 450$) and 4 ($n = 39$). Of these 489 students, 134 ($M_{age} = 8.71$ years, $SD = 0.87$, 61 girls [47.30%], 68 boys [50.70%]) were instructed by IG-teachers and 355 ($M_{age} = 8.40$ years, $SD = 0.56$, 179 girls [51.40%], 169 boys [48.60%]) were instructed by (waiting) CG-teachers. As cross-year teaching was only implemented in Berlin, all Grade 4 students formed part of the IG. Based on data from classroom-level aggregates, IG-students came from lower-SES families than CG-students ($t(10.66) = -1.66$, $p = .126$, $d = -0.82$), but there were no pronounced differences regarding the share of multilingual students ($t(24) = -0.90$, $p = .378$, $d = -0.38$). CG-students outperformed IG-students in prior knowledge on "floating and sinking" ($t(24) = -1.56$, $p = .133$, $d = -0.64$) but not on "evaporation and condensation" ($t(24) = -0.64$, $p = .53$, $d = -0.26$). Group differences on the language-related measures were negligible (science vocabulary: $t[24] = -0.57$, $p = .571$, $d = -0.24$; general academic vocabulary: $t[24] = -0.54$, $p = .598$, $d = -0.24$).

4.3. Assessment of the Amount and Quality of Teachers' Language Input During Classroom Instruction

We used transcripts of a total of 50 videotaped lessons, 25 on "floating and sinking" ($n_{IG} = 8$, $n_{CG} = 17$) and 25 on "evaporation and condensation" ($n_{IG} = 8$, $n_{CG} = 17$), as a basis for assessing the amount and quality of teachers' language input[3]. The transcripts cover 45 minutes of each double lesson, focusing on whole-classroom interaction. In particular, these were the introductory sequence (activation of prior knowledge), the instructional sequence (presentation and explanation of experiments for group work), and the reflection sequence at the end of the lesson (discussion and explanation of observations).

Teacher utterances were subsequently automatically coded regarding a variety of linguistic complexity features using the software *Common Text Analysis Platform* (CTAP; Chen & Meurers, 2016; Weiss & Meurers, 2022), a fully web-based, freely available and extensively researched platform for linguistic complexity analyses. Built on a Java Framework (cf. Chen & Meurers, 2016 for a detailed technical description of the architecture), CTAP allows for

3 Not all of the 27 teachers participated in the video-recordings of both double lessons. This means that the teacher samples for the two science topics are not exactly identical.

broad linguistic complexity modeling using a Natural Language Processing (NLP) pipeline that segments texts into sentences, tokens, characters, and syllables and calculates various theory-based measures of linguistic complexity. This automated extraction process allowes to capture a large set of measures across different linguistic dimensions, including lexicon, syntax, morphology, discourse, language use and human processing – areas where manual analyses are often limited due to the high cost of implementation. Amongst others, CTAP has been used to characterize academic language development (Weiss & Meurers, 2019) and is currently the most extensive feature set for German complexity assessment. Of the 543 linguistic complexity features currently implemented in CTAP, we selected 18 features that matched the linguistic focus of our intervention. As shown in detail in Table 1, the majority of features considered in our analyses referred to teachers' vocabulary use (e.g., lexical diversity, lexical sophistication, morphological complexity), given the lesson focus on the introduction and use of new concepts and adequate terminology. Both videotaped lessons included active experimentation and the formulation and justification of hypotheses. With regard to syntactical features, we therefore considered teachers' use of causal connectives and, as these should lead to longer sentences, average sentence length.

4.4. Analytical Procedure

We used *t*-tests for comparing IC-CG-differences in teachers' oral language use. While normal distribution within samples is typically considered a precondition for conducting parametric tests, *t*-tests have been shown to be robust against violations of the normality assumption (e.g., Kubinger et al., 2009). In the present sample, almost all features were normally distributed within IG and CG for both topics. Within-group homoscedasticity was given for all features for Topic 1 and for all but two features for Topic 2. In these cases (i.e., number of tokens, lexical diversity as measured by the HDD), we report results for the *Welch*-test. Additional analyses based on the nonparametric Mann-Whitney-U-Test yielded very similar result patterns as those presented below. We used the effect size *d* for evaluating the practical relevance of our findings, interpreting 0.20 as a small, 0.50 as a medium, and 0.80 as a large effect (Cohen, 1988).

Teachers might adapt their oral language input to their students' prior knowledge and language proficiency. Therefore, we additionally investigated correlations (not displayed here) between the 18 features on the amount and complexity of teachers' language input in both topics and various characteristics of classroom composition (i.e., prior knowledge on the topics "floating and sinking" and "evaporation and condensation", mastery of science vocabulary and general academic vocabulary, share of multilingual students in class, classroom average of family SES). Yet, hardly any significant correlations emerged and there was no systematic pattern across topics. As there was, thus, no clear indication that teachers' language input covaried with their students' prior knowledge or the sociodemographic classroom composition, we refrained from controlling for any of these features in subsequent analyses. Moreover, students' oral language use and related features (e.g., length of student-teacher interactions in whole-classroom interactions) were not considered in the present analyses. However, most of the selected features are not directly impacted by the length of the analyzed utterances (see Table 1).

Table 1: Linguistic Features and Their Alignment With the Intervention

#	Feature Set	Description and Selected Features	Alignment With the Intervention
1	Surface features	Overall amount of words and sentences spoken by the teachers, i.e., (1) number of tokens (total number of words) and (2) number of types (number of unique words), (3) number of sentences, (4) average sentence length	Increasing the amount of (high-quality) language input by using strategies such as thinking-aloud techniques or mapping actions with language, acting as a language role model by using adequate sentence structures for hypothesizing and justifying
2	Lexical diversity	Variability of teachers' vocabulary as measured by indicators that most efficiently control for average text length, i.e., (5) *Hypergeometric Distribution Diversity* (HD-D; McCarthy & Jarvis, 2010) and (6) *Measure of Textual Lexical Diversity* (MTLD; McCarthy & Jarvis, 2010)	Definition, use, and consolidation of domain-specific vocabulary such as "to displace", "water cycle", or "gaseous"
3	Lexical density	Share of selected parts of speech, e.g., interrogative pronouns, nouns, and verbs, per overall tokens in the text (7–12)	As language role models, teachers should increasingly use academic language which is characterized by a high information density (e.g., Heppt & Schröter, 2023); asking language-supportive questions was one of the core language-support strategies targeted in the PD and should come along with a higher density of interrogative pronouns
4	Lexical elaboration	Lexical sophistication as reflected in age of active use (with more sophisticated vocabulary being actively used at an older age). Measures were extracted based on the corpus *Karlsruhe Children's Texts* (13–14; Lavalley et al., 2015; Weiss, 2023). While such corpora can be used for determining the age at which specific words are actively used, we use them for assessing the overall lexical sophistication of a teacher's vocabulary without identifying the exact words the teacher used	Definition, use, and consolidation of domain-specific vocabulary, which is likely to occur infrequently in everyday conversations and should, thus, be primarily acquired in instructional settings (Heppt & Schröter, 2023)
5	Cohesion (discourse)	Use of causal connectives, measured by (15) the overall number of connectives and (16) the number of connectives per token	Acting as a language role model in formulating justifications and explanations
6	Morphological complexity	Measures assessing (17) the share of tokens and (18) the share of types with at least three syllables	Definition, use, and consolidation of domain-specific vocabulary, which frequently consists of compounds or derivations, resulting in relatively long and morphologically complex words (Bailey et al., 2007; Köhne et al., 2015)

5. Results

We examined group differences in IG and CG teachers' amount and quality of language input during science teaching separately for Topic 1 ("floating and sinking") and Topic 2 ("evaporation and condensation"). The descriptive statistics, the statistics of the *t*-tests as well as the effect size *d* for evaluating the practical relevance of the effects are displayed in Tables 2 and 3.

For "floating and sinking", findings indicate that, overall, IG-teachers provided more language input than CG-teachers, resulting in large effect sizes (see Table 2). However, contrary to our expectations, CG teachers produced longer sentences than teachers from the IG. Regarding the selected features of lexical complexity, only sporadic group differences emerged, some of them in favor of the IG, others in favor of the CG. Specifically, teachers from IG and CG did not differ in the lexical diversity used in oral classroom talk. While we observed a higher density of verbs in the IG and a higher density of adverbs in the CG, no differences emerged for any of the other density features, including interrogative pronouns and content words. There is, thus, no clear indication that teachers' classroom talk was characterized by a higher information density in the IG than in the CG. Indicators used for assessing lexical elaboration do not point to systematic differences between IG and CG either.

In line with our assumptions, IG-teachers used more causal connectives. This large effect can mainly be explained by the overall larger amount of teacher talk in the IG, as the group differences do not persist when considering the number of tokens. Finally, compared to the CG, IG-teachers' classroom talk was characterized by a higher morphological complexity, as reflected in the higher share of tokens and types with at least three syllables.

In a next step, we compared these findings with the results for Topic 2 with the aim of exploring whether group differences in oral language input persist across topics. Investigating the results for Topic 2 revealed three following major findings (Table 3). First, when teaching the double lesson on "evaporation and condensation", IG and CG teachers did not differ in the overall amount of speech delivered. This finding is due to an increase of language input in the CG (difference in the number of tokens: $t(14) = -4.57$, $p < .001$, $d = 1.33$). In line with this finding, no group differences emerged in the number of causal connectives used in oral classroom talk. Second, the pattern of results remained largely stable for lexical complexity. Third, and in line with the findings for Topic 1, IG teachers used a higher share of morphologically more complex (i.e., longer) words.

Table 2: Descriptive and Inferential Statistics for Comparisons Between IG (n = 8) and CG (n = 17) in Language Input During Topic 1

Feature Set	#	Feature	IG		CG						Evaluation of IG-CG-differences
			M	*SD*	*M*	*SD*	*t*	*df*	*p*	*d*	
Surface Features	1	Number of tokens	3299.38	567.34	2455.29	471.69	3.92	23	<.001***	**1.68**	**IG > CG**
	2	Number of types	685.25	101.48	566.65	82.50	3.12	23	.002**	**1.34**	
	3	Number of sentences	602.25	152.83	431.18	139.73	2.77	23	.005**	**1.19**	
	4	Average sentence length	6.17	0.26	6.35	0.44	-1.05	23	.153	*-0.46*	IG < CG
Lexical diversity	5	HD-D	0.85	0.01	0.85	0.01	0.28	23	.392	0	IG = CG
	6	MTLD	58.29	7.39	58.15	7.96	0.04	23	.484	0.02	
Lexcial density	7	Interrogative pronouns	0.03	0.01	0.03	0.01	1.17	23	.126	0	no systematic IG-CG-differences
	8	Adjectives	0.05	0.01	0.05	0.01	0.27	23	.359	0	
	9	Adverbs	0.16	0.02	0.18	0.02	-2.21	23	.019*	*-1.00*	
	10	Content words	0.51	0.02	0.51	0.02	-0.69	23	.248	0	
	11	Nouns	0.15	0.01	0.15	0.02	0.35	23	.365	0	
	12	Verbs	0.22	0.01	0.21	0.01	2.50	23	.010*	**1.00**	
Lexical elaboration	13	Mean age at active use (content words, tokens)	10.33	0.07	10.35	0.09	-0.61	23	.275	-0.24	IG = CG
	14	Mean age at active use (content words, types)	10.38	0.06	10.38	0.08	-0.11	23	.454	0	
Cohesion (dis-course)	15	Number of causal connectives	56.63	17.02	37.82	14.93	2.81	23	.005**	**1.21**	**IG > CG**
	16	Number of causal connectives per token	0.02	0.004	0.02	0.01	0.84	23	.205	0	
Morphological complexity	17	Share of tokens with at least 3 syllables	0.36	0.02	0.35	0.02	1.10	23	.141	**0.50**	**IG > CG**
	18	Share of types with at least three syllables	0.61	0.03	0.58	0.03	2.16	23	.021	**1.00**	

Note. Group differences of at least medium effect size in favor of the IG **are printed in bold**. Group differences of at least medium effect size in favor of the CG *are printed in italics*. HD-D = Hypergeometric Distribution Diversity; MTLD = Measure of Textual Lexical Diversity.
$*p < .05$. $** p < .01$. $*** p < .001$.

Table 3: Descriptive and Inferential Statistics for Comparisons Between IG (n = 8) and CG (n = 17) in Language Input during Topic 2

Feature Set	#	Feature	IG		CG						Evaluation of IG-CG-differences
			M	*SD*	*M*	*SD*	*t*	*df*	*p*	*d*	
Surface features	1	Number of tokens	3124.75	390.44	3196.35	694.18	-0.33	22.01	.373	-0.12	
	2	Number of types	713.25	69.38	689.00	105.77	0.59	23	.281	0.25	IG = CG
	3	Number of sentences	463.25	107.36	464.18	120.65	-0.02	23	.493	-0.01	
	4	Average sentence length	6.49	0.46	6.60	0.38	-0.60	23	.279	-0.27	IG = CG
Lexical diversity	5	HD-D	0.86	0.01	0.86	0.01	0.09	8.91	.466	0	IG = CG
	6	MTLD	63.40	11.54	62.63	7.19	0.20	23	.421	0.09	
Lexcial density	7	Interrogative pronouns	0.03	0.01	0.03	0.01	-0.01	23	.496	0	no systematic IG-CG-differences
	8	Adjectives	0.06	0.01	0.06	0.01	-0.23	23	.410	0	
	9	Adverbs	0.17	0.02	0.18	0.02	-0.78	23	.221	*-0.50*	
	10	Content words	0.52	0.02	0.51	0.02	0.84	23	.204	**0.50**	
	11	Nouns	0.14	0.01	0.14	0.02	0.81	23	.212	0	
	12	Verbs	0.22	0.01	0.21	0.01	2.37	23	.013*	**1.00**	
Lexical elaboration	13	Mean age at active use (content words, tokens)	10.35	0.07	10.35	0.07	0.01	23	.498	0	IG = CG
	14	Mean age at active use (content words, types)	10.44	0.11	10.41	0.09	0.63	23	.269	0.31	
Cohesion (dis-course)	15	Number of causal connectives	49.25	20.18	49.53	15.10	-0.04	23	.485	-0.02	IG = CG
	16	Number of causal connectives per token	0.02	0.01	0.02	0.01	-0.08	23	.468	0	
Morphological complexity	17	Share of tokens with at least 3 syllables	0.36	0.03	0.34	0.02	1.64	23	.057†	**0.85**	**IG > CG**
	18	Share of types with at least three syllables	0.61	0.03	0.59	0.04	1.16	23	.129	**0.54**	

Note. Group differences of at least medium effect size in favor of the IG **are printed in bold**. Group differences of at least medium effect size in favor of the CG *are printed in italics*.
† $p < .10$. * $p < .05$.

6. Discussion

Delivering classroom instruction that benefits all students and that fosters both conceptual understanding and language proficiency is highly demanding. There is, thus, a need for effective PD for helping teachers acquire the necessary skills. Drawing on a PD approach that has been shown to promote teachers' knowledge on subject-integrated language-support and benefits their language-supportive classroom teaching (Gabler et al., 2024; Heppt et al., 2022), the present study investigated teachers' oral language use in regular elementary school science classrooms. Specifically, using CTAP (Chen & Meurers, 2016; Weiss & Meurers, 2022), we investigated (1) whether teachers who had participated in a PD program on language-supportive teaching provided their students with quantitatively and qualitatively more stimulating language input than did teachers from the (waiting) CG and (2) whether these effects were sustained over time. For Topic 1 on "floating and sinking", we found that IG teachers talked more than CG teachers during classroom teaching and that they used more causal connectives. Moreover, their language input was characterized by higher morphological complexity, as indicated by the share of words with more than three syllables, and this effect persisted irrespective of the overall amount of language input. For the other features of linguistic complexity, however, there were either no group differences (i.e., lexical diversity) or the result pattern was very heterogeneous and did not allow for clear conclusions (i.e., lexical density and lexical elaboration). The result pattern for Topic 2 on "evaporation and condensation" was similar, except for the overall amount of language input. As CG teachers increased the amount of language input from Topic 1 to Topic 2, no group differences emerged in the amount of language input and the use of causal connectives. Overall, considering both topics, only a small number of features point to more sophisticated and complex language use in the IG.

The study adds to the small body of research focusing on teachers' amount and sophistication of language input when investigating the effectiveness of PD for language-supportive teaching. Prior research reported strong positive effects of a 3-hour PD on teachers' lexical diversity, measured in the number of word types, and the use of general and domain-specific academic vocabulary (Henrichs & Leseman, 2014). This study used very specific definitions of what was rated as general and specific vocabulary based on detailed word lists. Another intervention study that was conducted over the course of several months and included video feedback sessions, increased teachers' use of syntactically complex sentences and their lexical sophistication, as indicated in the overall word frequency (van Dijk et al., 2019). In the present study, IG and CG classroom talk differed in just a few linguistic features. Yet, whereas previous studies typically considered only very few language features, we included six feature sets with a total of 18 features.

The finding that IG teachers provided more language input during Topic 1, along with a more frequent use of causal connectives and an enhanced use of morphologically complex words, is basically in line with our assumptions. IG teachers were trained for the use of the language-support strategy language modeling, incorporating an increase of language input by thinking-aloud techniques and mapping their own or their students' actions with language. Acting as language role models, IG teachers deliberately aimed at performing important language functions of scientific inquiry, such as justifying their hypotheses and conclusions, which requires the use of causal connectives (e.g., Gabler et al., 2020). Their more frequent use of long and morphologically complex words, which persisted across both topics, matches the nature of domain-specific and general academic vocabulary. Such terms have repeatedly been identified to be relatively long (Bailey et al., 2007), in particular in the German language, where the use of compounds and verbs with prefixes is quite common (Köhne et al., 2015).

In explaining the overall small amount of systematic IG-CG-differences, which were still more pronounced in Topic 1 than in Topic 2, several aspects need to be considered. First, the PD on subject-integrated language support drew on the lesson plans and teaching materials of Topic 1. It required teachers to teach at least two lessons on "floating and sinking" in their Grade 3 classrooms, while actively implementing core language-support strategies such as language modeling. The implementation of language-supportive behavior was further supported by detailed suggestions for language support that were included in the lesson plans for Topic 1 in the IG. Feedback and video-coaching sessions were also based on teachers' experiences while teaching the topic "floating and sinking". IG teachers, thus, had ample opportunity to familiarize with Topic 1 and to plan and use language-support strategies while teaching this topic. This probably helped them to integrate strategies such as thinking-aloud techniques or the use of core domain-specific vocabulary in different contexts and contrastive ways, resulting in an enhanced language input during Topic 1. Topic 2 required IG teachers to transfer their knowledge and skills without prior classroom trials and without receiving suggestions for language support in the lesson plans, possibly resulting in reduced IG-CG-differences (Heppt et al., 2022).

Second, with the aim of ensuring a high level of comparability across intervention groups, teachers from both groups taught exactly the same lessons, using detailed lesson plans and accompanying teaching materials. It can therefore be assumed that teachers' language use was, at least to a certain degree, predetermined by the lesson content. In particular, academic vocabulary needed for co-constructing knowledge on core concepts and for verbalizing assumptions and observations, occurred in lesson plans and work sheets and had to be used by teachers in IG and CG alike. In a similar line of reasoning, it should be noted that language use across topics surely does not only depend on PD participation but also on differences that are inherent to the topics. Thus, vocabulary like "metal", or "plastic", that plays an important role for "floating and sinking", is more frequent and, thus, less elaborate than vocabulary such as "to rise", or "teapot warmer", that occurs in the lesson unit on "evaporation and condensation".

Third, teachers from both groups participated in PD courses on the two elementary school science topics, suggesting that both IG- and (waiting) CG-teachers were well-prepared for providing high-quality science instruction and were motivated to engage in teacher PD. As previous research demonstrated a positive link between certain teacher characteristics, such as occupational motivation, constructivist beliefs, and content knowledge, and the uptake of teacher PD (Richter et al., 2021), it is possible that teachers of both groups had better prerequisites for delivering cognitively activating and language-supportive instruction than the average teacher has. Possibly, group differences in language use would have been more pronounced if CG teachers had delivered the elementary school science classes based on the curricula and teaching materials only, without participating in respective PD courses.

6.1. Limitations and Conclusion

A major limitation of our study is that it is based on only a small convenience sample of teachers from two German federal states. This resulted in limited test power and prevents generalization of results to the larger population of elementary school teachers. Moreover, data on teachers' instructional practice and, hence, their oral language use in classroom instruction have only been collected after PD completion. Consequently, it is not possible for us to investigate changes in the quantity and quality of teachers' language input after participating in our PD program. Yet, previous analyses with the same sample showed that IG-teachers' knowledge on subject-integrated language support increased after PD completion and that they outperformed their counterparts from the CG (Heppt et al., 2022). It therefore seems reasonable to assume that differences in teachers' classroom talk are, at least partly, attributable to different treatment conditions.

The linguistic complexity features used for comparing the quasi-experimental groups were chosen with great care to yield maximum alignment with the intervention. Although word length is a widely established indicator of morphologic complexity (e.g., Bailey et al., 2007; Köhne et al., 2015), one could still argue that it is a rather simplistic feature. A more direct assessment of different forms of derivations, such as nominalizations or compounds, might prove even more insightful. Along similar lines, for gaining a better understanding of teachers' language-supportive instructional practice, it might be helpful to not only learn to what degree teachers use long words or causal connectives but rather which words they use and in which instructional contexts and phases of instruction.

Despite these shortcomings, we believe that using computational linguistic analysis methods for investigating PD effects offers valuable insights. In particular, the results of the present study complement and validate prior findings on the same PD approach, pinpointing that differences between the intervention groups are most pronounced with Topic 1 and get attenuated with Topic 2 – a topic which was less familiar to IG teachers in terms of material and respective linguistic scaffolding. Using computational linguistic analysis methods also expands on prior findings. Whereas IG teachers' science instruction on "floating and sinking" has previously been shown to be of higher language-supportive quality, relying on a highly inferential global rating (Heppt et al., 2022), the fine-grained linguistic analyses yielded through computational linguistic analyses gives hints on the language features that might have contributed to these differences. In order to gain an even more comprehensive understanding of teachers' amount and complexity of oral language, it would be highly valuable for future research to also consider students' language use, ideally also including the number and length of teacher-student as well as teacher-whole classroom interactions. In our study, there were no measurable covariations between students' prior knowledge, indicated by performance on achievement tests, and teachers' language use. Considering student-teacher interactions would allow for further insights as to whether teachers adapt their language input to their students' knowledge and language proficiency, which is an important prerequisite for fostering learning.

References

Babinski, L. M., Amendum, S. J., Knotek, S. E., Sánchez, M., & Malone, P. (2018). Improving young English learners' language and literacy skills through teacher professional development: A randomized controlled trial. *American Educational Research Journal*, *55*(1), 117–143. https://doi.org/10.3102/0002831217732335

Bailey, A. L., Butler, F. A., Stevens, R., & Lord, C. (2007). Further specifying the language demands of school. In A. L. Bailey (Ed.), *The language demands of school. Putting academic English to the test* (pp. 103–156). Yale University Press.

Becker-Mrotzek, M., & Woerfel, T. (2020). Sprachsensibler Unterricht und Deutsch als Zweitsprache als Gegenstand der Lehrerbildung [Language-supportive teaching and German as a second language as a topic for teacher education]. In C. Cramer, J. König, M. Rothland, & S. Blömeke (Eds.), *Handbuch Lehrerinnen- und Lehrerbildung* (pp. 98–104). Klinkhardt. https://doi.org/10.35468/hblb2020-011

Bravo, M. A., & Cervetti, G. N. (2014). Attending to the language and literacy needs of English learners in science. *Equity & Excellence in Education*, *47*(2), 230–245. https://doi.org/10.1080/10665684.2014.900418

Chen, X., & Meurers, D. (2016). CTAP: A web-based tool supporting automatic complexity analysis. *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*, 113–119. COLING.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession*. National Staff Development Council. https://edpolicy.stanford.edu/sites/default/files/publications/professional-learning-learning-profession-status-report-teacher-development-us-and-abroad.pdf

Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development*, *82*(3), 870–886. https://doi.org/10.1111/j.1467-8624.2011.01576.x

Dokter, N., Aarts, R., Kurvers, J., Ros, A., & Kroon, S. (2017). Academic language in elementary school mathematics: Academicness of teacher input during whole class instruction. *Dutch Journal of Applied Linguistics*, *6*(2), 213–230. https://doi.org/10.1075/dujal.17007.dok

Ernst-Slavit, G., & Mason, M. R. (2011). "Words that hold us up:" Teacher talk and academic language in five upper elementary classrooms. *Linguistics and Education*, *22*(4), 430–440. https://doi.org/10.1016/j.linged.2011.04.004

Gabler, K., Henschel, S., Hardy, I., Sontag, C., & Heppt, B. (2024). Sprachförderliches Unterrichten im Sachunterricht der Grundschule [Language-supportive teaching in elementary school science]. *Zeitschrift für Grundschulforschung*, *17*, 87–110. https://doi.org/10.1007/s42278-024-00191-8

Gabler, K., Mannel, S., Hardy, I., Henschel, S., Heppt, B., Hettmannsperger-Lippolt, R., Sontag, C., & Stanat, P. (2020). Fachintegrierte Sprachförderung im Sachunterricht der Grundschule: Entwicklung, Erprobung und Evaluation eines Fortbildungskonzepts auf der Grundlage des Scaffolding-Ansatzes [Subject-integrated language support in elementary school subject classes: Development, testing, and evaluation of a training concept based on the scaffolding approach]. In C. Titz, S. Weber, H. Wagner, A. Ropeter, C. Geyer, & M. Hasselhorn (Eds.), *Sprach- und Schriftsprachförderung wirksam gestalten: Innovative Konzepte und Forschungsimpulse* (pp. 59–83). Kohlhammer.

Gibbons, P. (2002). *Scaffolding language, scaffolding learning. Teaching second language learners in the mainstream classroom*. Heinemann.

Hart, B., & Risely, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H. Brookes.

Henrichs, L. F., & Leseman, P. P. M. (2014). Early science instruction and academic language development can go hand in hand. The promising effects of a low-intensity teacher-focused intervention. *International Journal of Science Education*, *36*(17), 2978–2995. https://doi.org/10.1080/09500693.2014.948944

Henschel, S., & Heppt, B. (2024). *Kompetenzeinschätzungen von Grundschullehrkräften zum sprachbildenden Unterrichten im Fach Mathematik: Zusammenhänge mit Merkmalen der Aus- und Fortbildung, Unterrichtsaktivitäten und mathematischen Kompetenzen der Schüler:innen* [Primary school teachers' self-perceived competencies in language-supportive mathematics instruction: relationships with characteristics of teacher education and professional development, teaching activities, and students' mathematical competencies] [Manuscript submitted for publication]. Institut zur Qualitätsentwicklung im Bildungswesen (IQB).

Henschel, S., Heppt, B., Rjosk, C., & Weirich, S. (2022). Zuwanderungsbezogene Disparitäten [Immigration-related disparities]. In P. Stanat, S. Schipolowski, R. Schneider, K. Sachse, S. Weirich, & S. Henschel (Eds.), *IQB-Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich* (pp. 181–219). Waxmann.

Heppt, B., Henschel, S., Hardy, I., Hettmannsperger-Lippolt, R., Gabler, K., Sontag, C., Mannel, S., & Stanat, P. (2022). Professional development for language support in science classrooms: Evaluating effects for elementary school teachers. *Teaching and Teacher Education*, *109*, Article 103518. https://doi.org/10.1016/j.tate.2021.103518

Heppt, B., & Schröter, P. (2023). Bildungssprache als übergeordnetes Ziel sprachlicher Bildung [Academic language as superordinate goal of language education]. In M. Becker-Mrotzek, I. Gogolin, H.-J. Roth, & P. Stanat (Eds.), *Grundlagen der sprachlichen Bildung* (pp. 139–153). Waxmann.

Kalinowski, E., Egert, F., Gronostaj, A., & Vock, M. (2020). Professional development on fostering students' academic language proficiency across the curriculum – A meta-analysis of its impact on teachers' cognition and teaching practices. *Teaching and Teacher Education*, *88*, Article 102971. https://doi.org/10.1016/j.tate.2019.102971

Kane, C., Sandilos, L., Hammer, C. S., Komaroff, E., Bitetti, D., & López, L. (2023). Teacher language quality in preschool classrooms: Examining associations with DLLs' oral language skills. *Early Childhood Research Quarterly*, *63*, 352–361. https://doi.org/10.1016/j.ecresq.2023.01.006

Köhne, J., Kronenwerth, S., Redder, A., Schuth, E., & Weinert, S. (2015). Bildungssprachlicher Wortschatz – linguistische und psychologische Fundierung und Itementwicklung [Academic vocabulary – linguistic and psychological foundation and item development]. In A. Redder, J. Naumann, & R. Tracy (Eds.), *Forschungsinitiative Sprachdiagnostik und Sprachförderung (FiSS) – Ergebnisse* (pp. 67–92). Waxmann.

Kubinger, K. D., Rasch, D., & Moder, K. (2009). Zur Legende der Voraussetzungen des t -Tests für unabhängige Stichproben [On the legend of the preconditions of the t-test for independent samples]. *Psychologische Rundschau, 60*(1), 26–27. https://doi.org/10.1026/0033-3042.60.1.26

Lavalley, R., Berkling, K., & Stüker, S. (2015). Preparing children's writing database for automated processing. *Proceedings of the L1 Teaching, Learning and Technology* (pp. 9–15). https://isl.iar.kit.edu/pdf/Lavalley2015.pdf

Lipowsky, F., & Rzejak, D. (2015). Key features of effective professional development programmes for teachers. *RICERCAZIONE*, *7*(2), 27–51.

Llosa, L., Lee, O., Jiang, F., Haas, A., O'Connor, C., Van Booven, C. D., & Kieffer, M. J. (2016). Impact of a large-scale science intervention focused on English language learners. *American Educational Research Journal*, *53*(2), 395–424. https://doi.org/10.3102/0002831216637348

Ludewig, U., Schlitter, T., Lorenz, R., Kleinkorres, R., Schaufelberger, R., Frey, A., & McElvany, N. (2022). *Die COVID-19 Pandemie und Lesekompetenz von Viertklässler*innen. Ergebnisse der IFS-Schulpanelstudie 2016–2021* [The COVID-19 pandemic and reading achievement in Grade 4. Results of the IFS school panel study 2016–2021]. https://ifs.ep.tu-dortmund.de/storages/ifs-ep/r/Downloads_allgemein/IFS_Schulpanelstudie.pdf [access on Januar 3, 2023]

Mahan, K. R. (2020). The comprehending teacher: Scaffolding in content and language integrated learning (CLIL). *The Language Learning Journal*, *50*(1), 74–88. https://doi.org/10.1080/09571736.2019.1705879
McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381
Paetsch, J., & Heppt, B. (2021). Teaching German as a second language: Implications for teacher education in Germany. In J. Chin-Kin Lee & T. Ehmke (Eds.), *Quality in teacher education and professional development: Issues and prospects in Chinese and German systems* (pp. 253–267). Routledge.
Piwowar, V., Thiel, F., & Ophardt, D. (2013). Training inservice teachers' competencies in classroom management. A quasi-experimental study with teachers of secondary schools. *Teaching and Teacher Education*, *30*, 1–12. https://doi.org/10.1016/j.tate.2012.09.007
Prediger, S., & Hardy, I. (2023). Fachliches und sprachliches Lernen im mathematisch-naturwissenschaftlichen Unterricht [Content- and language-related learning in mathematics and science instruction]. In M. Becker-Mrotzek, I. Gogolin, H.-J. Roth, & P. Stanat (Eds.), *Grundlagen der sprachlichen Bildung* (pp. 171–184). Waxmann.
Prediger, S., & Neugebauer, P. (2021). Can students with different language backgrounds profit equally from a language-responsive instructional approach for percentages? Differential effectiveness in a field trial. *Mathematical thinking and learning*, *25*(1), 2–22. https://doi.org/10.1080/10986065.2021.1919817
Prediger, S., & Wessel, L. (2017). Brauchen mehrsprachige Jugendliche eine andere fach- und sprachintegrierte Förderung als einsprachige? [Do multilingual students require other kinds of content- and language integrated interventions than monolingual students?]. *Zeitschrift für Erziehungswissenschaft*, *21*(2), 361–382. https://doi.org/10.1007/s11618-017-0785-8
Richter, E., Kunter, M., Marx, A. & Richter, D. (2021). Who participates in content-focused teacher professional development? Evidence from a large scale study. *Frontiers in Education, 6*, Article 722169. https://doi.org/10.3389/feduc.2021.722169
Sachse, K. A., Jindra, C., Schumann, K., & Schipolowski, S. (2022). Soziale Disparitäten [Social disparities]. In P. Stanat, S. Schipolowski, R. Schneider, K. Sachse, S. Weirich, & S. Henschel (Eds.), *IQB-Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich* (pp. 151–180). Waxmann.
Studhalter, U. T., Leuchter, M., Tettenborn, A., Elmer, A., Edelsbrunner, P. A., & Saalbach, H. (2021). Early science learning: The effects of teacher talk. *Learning and Instruction*, *71*, Article 101371. https://doi.org/10.1016/j.learninstruc.2020.101371
van Dijk, M., Menninga, A., Steenbeek, H., & van Geert, P. (2019). Improving language use in early elementary science lessons by using a video feedback intervention for teachers. *Educational Research and Evaluation*, *25*(5–6), 299–322. https://doi.org/10.1080/13803611.2020.1734472
Volodina, A., Heppt, B., & Weinert, S. (2021). Effects of socioeconomic status and language use on academic language proficiency in children with a migration background: An evaluation using quantile regressions. *Contemporary Educational Psychology, 65*, Article 101973. https://doi.org/10.1016/j.cedpsych.2021.101973
Vorholzer, A., & Aufschnaiter, C. (2019). Guidance in inquiry-based instruction – An attempt to disentangle a manifold construct. *International Journal of Science Education*, *41*(11), 1562–1577. https://doi.org/10.1080/09500693.2019.1616124
Weiss, Z. (2023). *An integrative approach to linguistic complexity analysis for German* [Doctoral dissertation, University of Tübingen]. Univerity Library Publication System Tübingen. https://doi.org/10.15496/publikation-93806

Weiss, Z., Lange-Schubert, K., Geist, B., & Meurers, D. (2022). Sprachliche Komplexität im Unterricht [Linguistic complexity in the classroom]. *Zeitschrift für Germanistische Linguistik*, *50*(1), 159–201. https://doi.org/10.1515/zgl-2022-2052

Weiss, Z., & Meurers, D. (2019). Analyzing linguistic complexity and accuracy in academic language development of German across elementary and secondary school. *Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 380–393). https://doi.org/10.18653/v1/W19-4440

Weiss, Z., & Meurers, D. (2022). Assessing sentence readability for German language learners with broad linguistic modeling or readability formulas: When do linguistic insights make a difference? *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, 141–153. Association for Computational Linguistics.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x