Instructional support in inquiry-based elementary school science classes: how does it relate to students' science content knowledge and academic language proficiency?

Birgit Heppt[1] · Sofie Henschel[2] · Ilonca Hardy[3] · Katrin Gabler[4]

Abstract

Inquiry-based science instruction has been proposed as an optimal learning environment for language-integrated teaching. While its potential for developing both content knowledge and language skills has been shown for students with limited language proficiency, research focusing on mainstream classrooms has mainly considered domain-specific learning. Despite the effectiveness of inquiry-based science instruction for student outcomes, research on the role of specific aspects of instructional quality is limited. Addressing this research gap, the present study investigates the relationship between teachers' instructional support during inquiry-based science classes and elementary school students' learning gains in science content knowledge and academic language proficiency. Multilevel regression analyses are based on data from 459 German elementary school students from Grades 3 and 4 who participated in a longitudinal intervention study that took place over one school year. Our findings indicate a strong impact of students' prior knowledge on learning outcomes and compositional effects for the language-related measures. Relations between teachers' instructional support, as measured by the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008), and students' learning outcomes did not emerge. In interpreting the results, the overall mediocre quality of teachers' instructional support needs to be considered.

Keywords Elementary school · Science education · Academic language · CLASS · Instructional support

---

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

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

2   Institute for Educational Quality Improvement (IQB), Humboldt-Universität zu Berlin, Berlin, Germany

3   Goethe University Frankfurt, Frankfurt, Germany

4   Freie Universität Berlin, Berlin, Germany

Introduction

It is generally accepted in educational research and instructional practice that academic achievement is related to students' language skills (e.g., Francis & Stephens, 2018; Golinkoff et al., 2018; Guo et al., 2015; Kempert et al., 2019). In particular, mastery of the language register of schooling, the so-called academic language register, has been identified as an important precondition of school success (Bailey, 2007; Meneses et al., 2018; Schuth et al., 2017; Snow, 2010). Yet, gaining academic language proficiency is a challenging task and many students need targeted support to develop the language skills needed for school success (e.g., Eckhardt, 2008; Francis & Stephens, 2018). Therefore, language-supportive teaching is widely considered a basic teaching principle for all teachers and across subjects and grade levels (Becker-Mrotzek & Roth, 2017; Bunch, 2013). This is also reflected in current curriculum development that aims at fostering both content knowledge and (academic) language proficiency in domain-specific teaching, for instance, in science classes (Becker-Mrotzek & Roth, 2017; Snow & Lawrence, 2011). Inquiry-based science classes are likely to offer ideal prerequisites for developing both conceptual understanding and language proficiency (e.g., Bravo & Cervetti, 2014; Lee et al., 2019). A growing number of curriculum intervention studies points to the effectiveness of inquiry-based science instruction for promoting English language learners (ELLs; for a meta-analysis, see Estrella et al., 2018). Research conducted with mainstream classrooms has, however, mainly investigated domain-specific learning gains, neglecting their potential for boosting students' (academic) language development (for a meta-analysis, see Furtak et al., 2012). Moreover, although specific teaching strategies are considered particularly beneficial for student learning, previous research has mostly focused on overall intervention effects without considering individual differences in teachers' instructional support and their effects on student outcomes. Addressing this research gap, the present study investigates the relationship between instructional support in inquiry-based elementary school science classes and students' learning gains, considering both their science content knowledge and academic language proficiency. We thereby attempt to clarify the role of instructional support for two central outcomes of instruction.

Promoting students' content knowledge and (academic) language skills in inquiry-based science education

Researchers have repeatedly advocated the incorporation of a linguistic focus and language-supportive teaching in science instruction (e.g., Bravo & Cervetti, 2014; Francis & Stephens, 2018; Henrichs & Leseman, 2014; Lee et al., 2019; Llosa et al., 2016; Ødegaard et al., 2014; van Dijk et al., 2019). In particular, inquiry-based science classes are considered an optimal learning environment for developing both conceptual understanding and language proficiency as they offer rich opportunities for using and developing language in meaningful contexts (e.g., Bravo & Cervetti, 2014; Estrella et al., 2018; Francis & Stephens, 2018; Heppt et al., 2022; Lee, 2020; Lee et al., 2019; Zwiep & Straits, 2013).

In such classes, students typically engage in the fundamental steps of scientific inquiry; i.e., they formulate hypotheses, plan and conduct experiments, and draw conclusions from the observed results (e.g., Furtak et al., 2012; Minner et al., 2010; Vorholzer &

Aufschnaiter, 2019). These activities are considered crucial for fostering students' conceptual understanding (e.g., Furtak et al., 2012), but they also provide ample opportunities for the use of academic language (e.g., Lee, 2020; Ødegaard et al., 2014; Seah & Silver, 2018). For formulating and justifying hypotheses, for instance, students need to know and apply both general academic vocabulary (i.e., vocabulary used across domains, such as "hypothesize," "determine") and specialized academic vocabulary (i.e., domain-specific technical terms, such as "buoyancy force," "evaporate"). Such terms are often abstract and polysemous, and many students are not familiar with them from their everyday lives (e.g., Ardasheva et al., 2019; Snow, 2010). Furthermore, students need to know clause connectives that are typically used for initiating justifications (e.g., "because," "therefore") and use them correctly for building long and syntactically complex sentences. In describing their observations and explaining their results, students as early as elementary school level are expected to take a relatively objective stance and report their ideas logically and coherently (cf. Schleppegrell, 2012). All these features are typical of the academic register (e.g., Bailey, 2007; Schleppegrell, 2001) and, specifically, of the language of science (e.g., Ardasheva et al., 2019; Francis & Stephens, 2018; Frändberg et al., 2013; Seah et al., 2014; Snow, 2010). The mastery of academic language substantially contributes to student achievement (Ardasheva et al., 2017; Meneses et al., 2018; Schuth et al., 2017), thus highlighting the interrelatedness of language and conceptual understanding (Francis & Stephens, 2018; Haug & Ødegaard, 2014; Snow, 2010).

Empirical findings on the effectiveness of inquiry-based science instruction for learning gains in science and language

When summarizing prior research on the effectiveness of inquiry-based science instruction, studies aimed at mainstream classrooms can be differentiated from studies focusing on students with limited language proficiency. Studies conducted in mainstream classrooms have mainly focused on students' domain-specific knowledge but have largely neglected potential effects on students' (academic) language development. These studies provided evidence that inquiry-based instruction can bolster students' conceptual understanding, highlighting selected teaching strategies that seem particularly beneficial (for a meta-analysis, see Lazonder & Harmsen, 2016; Minner et al., 2010). Specifically, it has been shown that science instruction benefits students' conceptual learning if teachers employ scaffolding (e.g., giving adaptive hints and prompts that are gradually reduced as students master the task themselves, adequate sequencing of the lesson contents; Hardy et al., 2006) and feedback techniques (e.g., providing motivating and informative feedback on students' current conceptual understanding, adaptive selection of tasks; Decristan et al., 2015). Moreover, more general aspects of teaching quality, such as effective classroom management, and the use of open-ended questions and challenging tasks, have been shown to boost elementary school students' conceptual understanding in inquiry-based instruction (Fauth et al., 2019).

Studies aimed at meeting the needs of students with limited language proficiency were mostly conducted in the USA. Their target group is ELLs who frequently grow up as dual language learners (DLLs). Overall, this research highlights the effectiveness of inquiry-based instruction for ELLs in developing their performance in science and language-related measures, such as reading comprehension and academic language (August et al., 2009, 2014; Bravo & Cervetti, 2014; Estrella et al., 2018; Lara-Alecio et al., 2012; Llosa et al., 2016; Maerten-Rivera et al., 2016; Tong et al., 2014). In addition to support strategies

aimed at developing students' reading comprehension (e.g., reading expository texts cou-
pled with reading strategies) and activating their multilingual resources (e.g., by having
students create bilingual glossaries for academic vocabulary), many of these interven-
tions include teaching strategies that are generally considered to benefit students' language
development. Such strategies are, for instance, the use of realia or visualizations as scaf-
folds, the implementation of ongoing classroom discussions, language modeling by provid-
ing rich and elaborate language input and multiple explanations, and adequate linguistic
feedback (e.g., by rephrasing students' answers; August et al., 2009; Lee & Buxton, 2013).

This line of research mostly focuses on the overall effectiveness of specific curricula.
While a few studies additionally investigated the degree to which teachers incorporated
such strategies into their classroom teaching (August et al., 2014; Garza et al., 2018; Tong
et al., 2018), they typically do not report how this impacted the effectiveness of the inter-
vention. In their literature review, Francis and Stephens (2018) thus conclude that there is
a lack of studies relating individual differences in teaching behavior with student outcomes
in the context of language-supportive science instruction.

Instructional support as a basis for student learning

Different approaches have been proposed for assessing individual differences in teaching
behavior, comprising both student ratings and classroom observations. Among the observa-
tion instruments, the Classroom Assessment Scoring System (CLASS; Pianta & Hamre, 2009;
Pianta et al., 2008) stands out as a widely used and internationally validated tool for assessing
interaction quality in instructional settings (e.g., Hu et al., 2016; Leyva et al., 2015; Pakarinen
et al., 2010; Stuck et al., 2016). Moreover, its domain on *instructional support*, comprising
"concept development," "quality of feedback," and "language modeling," refers to core teach-
ing strategies that are considered key for students' conceptual and/or language learning. Con-
cept development describes teaching strategies that activate students' prior knowledge and
engage them in higher-order thinking skills. Teachers who are rated high on the dimension of
concept development regularly engage their students in formulating and evaluating assump-
tions and conducting experiments during instruction, for example. Such activities should help
students develop conceptual understanding and are at the core of inquiry-based science classes
(Vorholzer & Aufschnaiter, 2019). Quality of feedback describes the use of adequate scaffolds
(i.e., additional information, follow-up questions for students to explain their reasoning) that
prompt students' thought processes and expand their learning (Pianta et al., 2008). Although
feedback strategies are also frequently used for bolstering students' language development
(e.g., by correcting or rephrasing and extending students' utterances with more sophisticated
or appropriate terms; Heppt et al., 2022; Lyster & Saito, 2010), the CLASS does not spe-
cifically address language-related feedback but covers feedback in a broader sense. To be very
clear, it mainly refers to feedback strategies aimed at promoting students' content learning and
conceptual understanding which may also impact their language skills (e.g., when students are
asked to elaborate on their reasoning and assumptions).

Finally, language modeling encompasses teaching behavior that focuses primarily on
students' language development, such as engaging students in frequent discussions and
peer conversations, asking open-ended questions that require elaborate answers, and pro-
viding them with sophisticated and varied language input (Pianta et al., 2008). These
strategies or linguistic prompts (linguistic scaffolds) are also frequently used within the
scaffolding approach of language-supportive teaching (Gibbons, 2002). Building on the

fundamental principles of Vygotsky's theory of social learning (Wood et al., 1976), this approach aims at helping students gradually expand their language skills from everyday language to a more formal register of academic language (Heppt et al., 2022; Lucero, 2014).

The CLASS, hence, offers a sound basis for assessing teaching behavior that is deemed to benefit students' conceptual understanding and language skills. While the CLASS dimension of instructional support in general (Allen et al., 2013) and specific teaching strategies, such as the use of feedback (Decristan et al., 2015; Wisniewski et al., 2020), have already been shown to promote students' domain-specific learning and conceptual understanding, less is known on its relation to students' academic language development. Findings on the relationship between the CLASS domain *instructional support* and children's language skills are primarily based on investigations from early childhood education and care (ECEC) programs, yielding mixed results. Thus, some studies reported positive effects of the CLASS dimension of instructional support on, for instance, vocabulary and preliteracy skills, such as phonological awareness and print knowledge, of preschoolers (Slot et al., 2018) and first-graders (Cadima et al., 2010). The bulk of studies, however, including a recent meta-analysis summarizing the results of 19 studies (Perlman et al., 2016), reported very small or even no relations between kindergarten teachers' instructional support and children's language skills (e.g., Bihler et al., 2018; Guerrero-Rosada et al., 2021; Kohl et al., 2019; Sabol et al., 2018). In interpreting these results, it needs to be taken into account that, overall, relatively low quality of instructional support was reported across studies (Perlman et al., 2016). Thus, kindergarten teachers' strategies aimed at fostering higher-order thinking and promoting children's language use did not meet the threshold necessary for boosting children's language learning (cf. Burchinal et al., 2010). Elementary school teachers, however, might place more emphasis on instructional support than kindergarten teachers because students' conceptual understanding and development is a core aim of instruction in school, and school language is a particular obstacle for many students. Teachers' language-related instructional support during inquiry-based science instruction might thus have substantial impact on (academic) language proficiency in elementary school.

The present study

This study investigates the role of teachers' instructional support for promoting students' conceptual understanding *and* academic language skills in standardized lessons of inquiry-based elementary school science. We investigate the effects of the three dimensions of instructional support (i.e., concept development, quality of feedback, and language modeling) as measured with the CLASS (Pianta et al., 2008) on gains of students' science content knowledge and academic language proficiency in a multilevel repeated measures design. Considering the teaching strategies subsumed within each of the three dimensions, we expect concept development and quality of feedback to be particularly beneficial for students' gains in science content knowledge, whereas language modeling should primarily contribute to students' academic language skills.

Previous research has repeatedly identified substantial relations between students' prior knowledge (Geary et al., 2017; Simonsmeier et al., 2021) and various sociodemographic variables (e.g., gender, language background, socioeconomic status; OECD, 2019; Rosén et al., 2022; Sirin, 2005) with their learning outcomes. Moreover, student achievement is related to characteristics of the classroom composition. Thus, students typically show larger learning gains when they are grouped with high-achieving students (Becker et al., 2022; Schmerse, 2021) or with students from families with high socioeconomic status (SES; Rjosk et al., 2014; van Ewijk & Sleegers, 2010). Classroom characteristics are, in

turn, associated with instructional processes (Fauth et al., 2021; Kuger et al., 2016; Rjosk et al., 2014). We, therefore, included students' prior knowledge and various sociodemographic characteristics (i.e., language background, socioeconomic and educational family background, gender) in our analyses. In considering variables at the classroom level, we also extend prior research on the effectiveness of language-supportive inquiry-based science instruction which typically did not consider compositional effects.

Method

General study information

This research is part of the project *ProSach* ("Professional development on content-focused language support in elementary school science instruction"; German: *Professionalisierungs-maßnahmen zur bedeutungsfokussierten Sprachförderung im Sachunterricht der Grundschule*) that aimed at evaluating a professional development (PD) program for elementary school teachers in Germany. The project was conducted in two German federal states (*Länder*) over two full school years. It consisted of a PD phase in Year 1 and an implementation phase in Year 2. In Year 1, teachers participated in PD for teaching selected elementary school science topics (both intervention group [IG] and control group [CG]) and language support in science classrooms along the lines of the scaffolding approach (Gibbons, 2002; IG only). In Year 2, teachers delivered the complete science units to their regular Grade 3 or 4 classrooms (for detailed descriptions of study design, PD contents, and findings regarding the effects on teachers' knowledge and classroom behavior, see Heppt et al., 2022). There were no intervention effects on student outcomes (i.e., students from the IG whose teachers participated in PD on elementary school science *and* language-supportive teaching did not show larger learning gains than students from the CG whose teachers participated in PD on elementary school science topics only). Given the large variance in teaching behavior within both groups, the present investigation does not differentiate between IG and CG in examining the effects of instructional support on students' learning gains.

Sample

The present analyses are based on the data of 459 elementary school students who participated in the implementation phase of the project, i.e., whose teachers delivered the lesson units on "floating and sinking" and "evaporation and condensation," and who took the accompanying pre- and posttests. Students were distributed across 27 classrooms from 13 schools. The majority of students were in Grade 3 ($n = 420$) but as one of the two federal states adopts cross-year learning in elementary school, the sample additionally comprised 39 Grade 4 students. At T1, students were 8.49 years old on average ($SD = 0.68$) and half of them were girls ($n = 230$; 50%). Based on their self-reports, 236 students (51%) were classified as DLLs; i.e., they speak at least one language other than German at home. Most of these students indicated to speak German and another language at home ($n = 191$; 81%), while only 45 students (19%) reported not speaking German in their families.

The 27 participating teachers were on average 42.20 years old ($SD = 7.44$) when entering the project and 22 of them were female. While almost all teachers had completed a university degree in elementary school education ($n = 23$; 89%), only eight (31%) had studied science as a school subject. The gender distribution and training background of the sample are both

typical of elementary school teachers in the participating states and Germany overall (OECD, 2020). Participation in the study was voluntary for schools, teachers, students, and parents (who also completed a questionnaire). We obtained informed consent from all participants.

Study design and procedure

The project included units on the elementary school science topics "floating and sinking" (Topic 1), "evaporation and condensation" (Topic 2), and "education for sustainable development" (Topic 3), all of which are part of the elementary school science curricula of the participating states. Due to sample attrition throughout the implementation phase, the present study includes only Topic 1 and Topic 2. The science topics comprised detailed lesson plans of six (Topic 1) or five (Topic 2) double lessons (90 min each), respectively, and teachers implemented them in their regular science teaching in the first six months of the school year 2017/18 (Figure ESM 1 in the Electronic Supplementary Material [ESM]). Whereas the curriculum on "floating and sinking" aimed at developing students' understanding of the concepts of density, water displacement, pressure, and buoyancy force (e.g., Kleickmann et al., 2016), the curriculum on "evaporation and condensation" focused on the hydrological cycle. Both science curricula were developed based on design principles of inquiry-based science education. They included a wide range of hands-on activities for students and actively engaged them in the core steps of scientific inquiry, thus aiming at developing their conceptual understanding (Fauth et al., 2019; Furtak et al., 2012). By frequently prompting students to express their assumptions and discuss their observations in small groups or in class, the teaching units also aimed at using and developing language skills in meaningful contexts. The curriculum on "floating and sinking" has been evaluated before with proven effectiveness in fostering elementary school students' conceptual knowledge (Decristan et al., 2015; Fauth et al., 2019). The curriculum on "condensation and evaporation" was designed accordingly and piloted in several classrooms (Lange-Schubert et al., 2017).

To ensure a high level of comparability across classrooms, teachers participated in two 5-h PD courses, one for each science topic, focusing on content knowledge and pedagogical content knowledge needed for teaching the curricula and familiarizing participants with the lesson plans of both science topics. They received detailed lesson plans for each lesson and the necessary teaching materials (e.g., worksheets, objects for demonstrating experiments, and for the students to conduct hands-on activities).

We videotaped the second double lesson (90 min) of both topics during the implementation phase and used the video recordings to investigate implementation fidelity and instructional support. As a further indicator for implementation fidelity, teachers completed documentation forms after each double lesson, indicating which of the mandatory lesson elements they had implemented and whether they had used any further teaching materials. Additionally, we administered written assignments to the students before (T1) and after (T2) the lesson unit on "floating and sinking," before (T2) and after (T3) the lesson unit on "evaporation and condensation," and by the end of the school year (T4; Figure ESM 1). Data collection took place in the classroom setting during regular lesson time and was conducted by trained student assistants. At T1, students completed a questionnaire on their gender, age, and language background. Basic information on students' socioeconomic and educational backgrounds was collected with a parent questionnaire.

Measures

Instructional support

We used the three CLASS dimensions concept development, quality of feedback, and language modeling for assessing teachers' instructional support (Pianta et al., 2008). All three dimensions capture teaching behavior and strategies that typically form part of inquiry-based science instruction and were also deliberately encouraged in our intervention. Indicators used for assessing concept development are, for instance, the implementation of experiments, the activation of prior knowledge, and the use of real-world applications. Quality of feedback is based on an assessment of, among other things, teachers' adequate use of scaffolds, such as giving additional information and having students explain their thinking. These strategies were regularly implemented in the lesson units (e.g., by asking students to justify their assumptions about the floating and sinking of certain objects and materials or to explain why ice melts). Language modeling refers to the use of open-ended questions, elaborate language, and self- and parallel talk as well as the facilitation of frequent conversations, for example. These language-support strategies were at the core of the PD for language support (Heppt et al., 2022).

As recommended by the CLASS protocol, ratings are based on 20-min video clips. Specifically, we selected the introductory sequences of the videotaped second double lesson of Topic 1 and Topic 2 for the CLASS ratings. The introductory sequences were typically conducted as classroom discussions (e.g., in circle time), with teachers providing impulses for conversations, referring students back to previous lessons, or demonstrating experiments, thus facilitating the observation of concept development, quality of feedback, and language modeling (as opposed to sequences in which experiments are set up, materials are removed, or students are taking notes). Licensed raters who were blind to our study goals rated each dimension on a 7-point scale (1–2: low quality; 3–5: average quality; 6–7: high quality). Raters double-coded each video (Topic 1: $n = 25$, Topic 2: $n = 25$[1]). The intraclass correlation coefficients (ICC) were satisfactory to very good for both topics (concept development: ICC $= 0.82/0.84$, quality of feedback: ICC $= 0.60/0.89$, language modeling: ICC $= 0.74/0.86$). In the case of divergent ratings, we used the mean score of both ratings for further analyses.

Science content knowledge on floating and sinking

For assessing students' science content knowledge on "floating and sinking," we constructed a test (10 tasks with 32 items) assembled of different empirically validated instruments on "floating and sinking" (e.g., Hardy et al., 2006) and administered it before (T1) and after (T2) the teaching unit on "floating and sinking." Our test version included two tasks with multiple-choice (MC), six tasks with forced-choice items (FC), one open-ended question and one task with items in graphical response format. The test assesses students' understanding of the concept of water displacement and the floating and sinking of objects. Given the polytomous and ordered scoring of the items (0–2), we scaled the test scores of both time points based on a Partial Credit Model (cf. Masters & Wright, 1997), using ConQuest 4.5.2 (Adams et al., 2015). We linked the data of both time points longitudinally based on the mean/mean method (Fischer et al., 2016) and used weighted likelihood estimates (WLE; Warm, 1989) as ability scores in the subsequent analyses. In line with previous studies, the internal consistency of the

---

[1] Although all 27 teachers delivered the lesson units to their classes, only 25 of them participated in the video-recordings.

scale was rather low, particularly in the pretest ($\alpha_{T1/T2} = 0.52/0.66$), reflecting the heterogeneous and limited prior knowledge base of elementary school students on this topic.

Science content knowledge on evaporation and condensation

We used a short version of a validated test instrument for assessing students' content knowledge on "evaporation and condensation" (e.g., Kleickmann et al., 2010) shortly before and after the corresponding unit (i.e., at T2 and T3). The test taps students' conceptual understanding of the aggregation states of water and their transition processes and mainly refers to phenomena that students should be familiar with from their everyday lives (e.g., wet dishes at the sink that are dry after a while). The constructed test consists of 8 tasks with 48 FC items. As for the measure on "floating and sinking," tasks were scored with 0, 1, and 2 and calibrated using a Partial Credit Model in ConQuest 4.5.2 (Adams et al., 2015). Again, we linked the data longitudinally based on the mean/mean-method and used the linked WLE scores for each participant for further analyses. The reliability of the scale was low ($\alpha_{T2/T3} = 0.58/0.60$), probably due to the high difficulty of the test.

Academic language proficiency

Science vocabulary The assessment of students' domain-specific academic language proficiency was based on a researcher-developed scale on science vocabulary. While the original scale included items for all three science topics that formed part of the larger project, only items that pertained to the topics "floating and sinking" ($n_{items} = 12$) and "evaporation and condensation" ($n_{items} = 9$) were included in the present analyses. As a two-dimensional model ($AIC_{pre/post} = 5045.46/3612.37$, $BIC_{pre/post} = 5141.38/3708.29$), differentiating between the two science topics, did not fit the data better than a unidimensional construct ($AIC_{pre/post} = 5041.04/3609.14$, $BIC_{pre/post} = 5128.96/3697.06$, $\Delta\chi^2_{pre/post} = 0.42/0.76$, $df_{pre/post} = 2$, $p_{pre/post} = 0.81/0.68$), we integrated all 21 items into a single scale of science vocabulary. The target words were selected from the lesson units and accompanying teaching materials (e.g., handouts) and were deemed crucial for gaining conceptual understanding (e.g., *to displace water*, *water cycle*).

The items were constructed in multiple choice-format and required students to select the correct word out of four to complete gapped sentences, label depicted processes, or find synonyms, for instance. In addition to the printed items in students' test booklets, items were read aloud to ensure that also students with limited reading proficiency could answer them. The scale was administered at T1 and T4, using a multi-matrix design in which each student answered only a subset of items (Kolen, 2006). Item responses were calibrated based on a 1 PL item response model in ConQuest 5 (Adams et al., 2020) and, subsequently, linked longitudinally based on the mean/mean method. We used the resulting WLE scores as person estimates for all further analyses. The reliability of the scale was low ($\alpha_{T1/T4} = 0.58/0.53$), possibly because it was fairly easy, covered two topics, and used a variety of item formats, thus resulting in substantial heterogeneity.

General academic vocabulary and comprehension of connectives We drew on a validated and standardized test instrument for assessing elementary school students' general (i.e., cross-subject) academic language proficiency in German (Heppt et al., 2020). Specifically, we used shortened versions (16 items per scale) of the measures on general academic vocabulary (*BiSpra-Word*) and comprehension of connectives (*BiSpra-Sentence*). *BiSpra-Word* taps the comprehension of words that are used across subjects for explaining instructions and processes, for instance (e.g., *structure*, *to indicate*). *BiSpra-Sentence* focuses on the comprehension of different types of connectives (e.g., temporal, concessive, modal) that are more frequently used in formal settings than in everyday interactions (e.g., *although*, *however*).

In both scales, students have to select the semantically and grammatically correct word out of three (*BiSpra-Word*) or four (BiSpra-Sentence) to complete a gapped sentence. The items are printed in the student booklets and presented auditorily from CD to mitigate possible confounding effects of students' reading comprehension. We administered both scales at T1 and T4. The reliability of the scales was satisfactory (*BiSpra-Word*: $\alpha_{T1/T4} = 0.67/0.71$, *BiSpra-Sentence*: $\alpha_{T1/T4} = 0.76/0.79$). Analyses are based on sum scores.

Control variables

We used the Highest International Socio-Economic Index (HISEI; Ganzeboom et al., 1992) as an indicator of the families' SES. The HISEI is an index of the highest occupational status of both parents. It ranges from 10 to 90 with higher values indicating occupations with higher SES. For assessing students' educational background, we asked the parents about their highest educational qualification and transformed it into the number of years of education (OECD, 2009). The so-called PARED ranges from 4 years (elementary school) to 18 years (doctoral degree) and we used the highest PARED of both parents in our analyses. In addition, we controlled for gender (0 = boy, 1 = girl) and language background (0 = German monolingual, 1 = DLL).

Statistical analyses

The amount of missing data ranged from 0% (grade level) to 29% (HISEI) with 5% (Topic 1) and 7% (Topic 2) for the CLASS variables. To handle these missing data, we applied multiple imputation procedures that considered the clustered data structure (Grund et al., 2017). The imputation model was based on a random-intercept model. It included all variables used in the subsequent analyses and selected auxiliary variables (e.g., grade level) that were substantially related to the study variables. Based on the R packages mice (van Buuren & Groothuis-Oudshoorn, 2011) and mitml (Grund et al., 2021), we generated 30 full datasets. For analyzing the effects of teachers' instructional support (Level 2) on students' science knowledge and academic language proficiency (Level 1), we subsequently conducted separate multilevel regression analyses for all dependent variables in M*plus* 8.4 (doubly manifest random-intercept models; Muthén & Muthén, 1998–2017). Metric Level 1 predictors were group-mean centered before the analyses. All Level 1 predictors were additionally entered into the models as classroom-aggregated and *z*-standardized variables at Level 2. The results of the 30 analyses were combined, using the option "type = imputation" (Rubin, 1987). Code files of the analyses in M*plus* can be found on OSF: https://osf.io/7xdkf/?view_only=dd1f3189da014cccb23a09c137ac3d7a.

Table 1 Descriptive statistics of all study variables

Variable		M/%	SD
Student-level variables			
1	Content knowledge "floating and sinking" (T1)[a]	0.01	0.85
2	Content knowledge "floating and sinking" (T2)[a]	0.40	1.00
3	Content knowledge "evaporation and condensation" (T2)[a]	− 2.05	1.22
4	Content knowledge "evaporation and condensation" (T3)[a]	0.25	1.27
5	Science vocabulary (T1)[a]	− .04	1.04
6	Science vocabulary (T4)[a]	.99	1.14
7	General academic vocabulary (T1)	9.52	3.15
8	General academic vocabulary (T4)	10.94	3.12
9	Comprehension of connectives (T1)	8.93	3.49
10	Comprehension of connectives (T4)	10.53	3.49
11	DLL[b]	51	–
12	SES	56.76	20.21
13	Parental education[c]	14.71	3.65
14	Girl[b]	50	–
Classroom-level variables			
15	Concept development (Topic 1)	3.03	0.99
16	Quality of feedback (Topic 1)	2.56	0.66
17	Language modeling (Topic 1)	3.32	0.91
18	Concept development (Topic 2)	4.37	0.91
19	Quality of feedback (Topic 2)	3.52	1.02
20	Language modeling (Topic 2)	3.95	1.05

$N_{students} = 459$, $N_{classrooms} = 27$

Analyses are based on 30 imputed datasets and were combined based on Rubin's (1987) formulas. [a]Performance measure is based on a WLE score (Warm, 1989). [b]For binary variables, the percentage of category "1" is reported instead of the mean. *DLL*, dual language learner; *SES*, socioeconomic status (as measured by the HISEI). [c]Parental education as indicated by the highest number of education years of both parents (OECD, 2009)

Results

Preliminary analyses and descriptive statistics

We checked teachers' implementation fidelity as a baseline for all further analyses. Based on the videos of both science topics (90 min per topic and classroom), we examined whether participants implemented obligatory elements of the lesson plans in their classroom teaching. On average, teachers delivered 90.44% ($SD = 10.23\%$) of the compulsory lesson elements of Topic 1 and 88.47% ($SD = 9.68$) of the compulsory lesson elements of Topic 2, indicating a highly satisfactory implementation fidelity (cf. Heppt et al., 2022).

Descriptive statistics of all study variables are shown in Table 1. Bivariate correlations of all variables at the classroom level (Table ESM 1) and at the student level (Table ESM 2) are presented in the ESM. With CLASS scores ranging from 2.56 ($SD = 0.66$) for quality of feedback in Topic 1 to 4.37 ($SD = 0.91$) for concept development in Topic 2, instructional support was mainly rated as being of average quality in the present sample (Table 1). Yet, overall, teachers showed significantly better instructional support for Topic 2 than

for Topic 1 (concept development: $t = 4.34$, $df = 26$, $p < 0.001$, $d = 1.41$; quality of feedback: $t = 3.97$, $df = 26$, $p < 0.001$, $d = 1.12$; language modeling: $t = 2.29$, $df = 26$, $p = 0.02$, $d = 0.64$). Results further indicate that students gained in content knowledge on "floating and sinking" from T1 to T2 ($t = 7.33$, $df = 458$, $p < 0.001$, $d = 0.42$) and on "evaporation and condensation" from T2 to T3 ($t = 30.37$, $df = 458$, $p < 0.001$, $d = 1.85$; Table 1) with substantially larger learning gains on the latter topic ($t = 20.92$, $df = 458$, $p < 0.001$, $d = 1.44$). Moreover, their achievement on the scales for science vocabulary ($t = 17.65$, $df = 458$, $p < 0.001$, $d = 0.94$), general academic vocabulary ($t = 6.89$, $df = 458$, $p < 0.001$, $d = 0.45$), and comprehension of connectives ($t = 8.08$, $df = 458$, $p < 0.001$, $d = 0.46$) improved across the school year. The learning gains on the science vocabulary scale were slightly larger than those on the two more general measures of academic language proficiency (general academic vocabulary: $t = 1.84$, $df = 458$, $p = 0.07$, $d = 0.16$; comprehension of connectives: $t = 3.06$, $df = 458$, $p < 0.05$, $d = 0.24$). No significant differences emerged in the increase in general academic vocabulary and comprehension of connectives ($t = 0.80$, $df = 458$, $p = 0.40$, $d = 0.07$).

While classroom-level aggregates of students' performance on the academic language measures were strongly correlated, smaller correlations emerged for students' achievement on the knowledge-related science measures. This might be due to the very different aspects of domain-specific knowledge that are covered by the measures of conceptual understanding of "floating and sinking" and "evaporation and condensation" (cf. Stadler et al., 2021). Correlations of the performance measures with concept development, quality of feedback, and language modeling were mostly negligible at both time points. The correlations between the three CLASS dimensions across topics were fairly small, thus pointing to low stability of instructional support over time. Hence, the small relations between instructional support and students' competencies might partly be due to the varying quality of instructional support.

Prediction of students' gains in science content knowledge and academic language comprehension

The results of the multilevel regression analyses are shown in Table 2 and in Table ESM 3. Table 2 focuses on the effects of instructional support in Topic 1 ("floating and sinking"); Table ESM 3 displays the findings for instructional support in Topic 2 ("evaporation and condensation"). The ICCs for content knowledge on "floating and sinking," science vocabulary, general academic vocabulary, and comprehension of connectives range from 0.13 to 0.17, pointing to substantial variance at the classroom level. However, for content knowledge on "evaporation and condensation," the ICC was very small (0.04). Posttest achievement of students within the same classroom were thus barely more strongly related than posttest achievement of students from different classrooms, resulting in limited potential of the classroom level variables for explaining variance.

Table 2 shows that several variables at the student level predicted knowledge on "floating and sinking" at T2, whereas none of the classroom level variables (i.e., neither instructional support nor classroom composition measures) was significantly related to students' posttest achievement (Model 1). Specifically, we found positive effects of prior knowledge, family SES, and parental education; yet, smaller learning gains occurred for DLLs than for monolingual students. A slightly different picture emerged for the language-related measures (Models 2–4). In line with the findings for science knowledge, students' individual prerequisites were significantly associated with science vocabulary, general academic vocabulary, and comprehension of connectives at T4. Again, prior knowledge turned out as

Table 2 Prediction of students' science content knowledge and academic language proficiency by instructional support during the lesson unit on "floating and sinking" (Topic 1)

	Science content knowledge ("floating and sinking")		Science vocabulary		General academic vocabulary		Comprehension of connectives	
	Model 1		Model 2		Model 3		Model 4	
	β	SE	β	SE	β	SE	β	SE
Level 1 (individual students)								
Prior knowledge	.33**	.04	.34**	.04	.37**	.06	.44**	.06
DLL	−.10*	.05	−.04	.05	−.13*	.05	−.05	.05
SES	.12*	.06	.07	.05	.04	.05	.11*	.06
Parental education[a]	.10†	.06	.04	.05	.07	.06	.05	.06
Girl	−.04	.05	.09†	.05	−.04	.04	.07†	.04
Level 2 (classrooms)								
Prior knowledge (mean)	.37	.23	.44*	.18	.47*	.20	.46†	.26
DLLs (%)	−.09	.19	−.53**	.19	−.05	.22	−.35	.22
SES (Mean)	.30	.47	.79*	.32	.23	.28	.30	.32
Parental education (mean)	.03	.42	−.35	.33	.34	.25	.06	.26
Girls (%)	.10	.28	−.19	.18	−.20	.15	−.30	.23
Concept development	−.14	.32	−.31	.22	−.23	.25	−.04	.21
Quality of feedback	.12	.31	.08	.20	−.22	.22	−.12	.27
Language modeling	.04	.34	.08	.23	.39	.26	−.09	.26
R^2 Level 1	.19**	.04	.15**	.03	.20**	.05	.26**	.05
R^2 Level 2	.35†	.18	.91**	.15	.85**	.11	.86**	.18
ICC	.14		.13		.15		.17	

$N = 459$. Analyses are based on 30 imputed datasets and were combined based on Rubin's (1987) formulas. Additionally, including the treatment condition in the analyses (0 = control group, 1 = intervention group) does not change the pattern of results. *DLL*, dual language learner; *SES*, socioeconomic status (as measured by the HISEI). [a]Parental education as indicated by the highest number of education years of both parents (OECD, 2009)

†$p < .10$. *$p < .05$. **$p < .01$

the most important predictor of all three measures. Being a DLL was negatively associated with students' learning gains in general academic vocabulary (Model 3) but not in science vocabulary and in the comprehension of connectives (Models 2 and 4). Girls showed larger learning gains on science vocabulary and comprehension of connectives (but not on general academic vocabulary). Moreover, family SES turned out as a significant predictor of comprehension of connectives at T4, indicating a stronger increase in the comprehension of connectives for students from high-SES families compared to students from socioeconomically disadvantaged families (Models 3 and 4). Again, none of the CLASS measures of instructional support was significantly related to students' learning gains. In addition—and different from the findings for science content knowledge—classroom composition had an impact on students' individual language development. For all three measures, larger learning gains occurred for students in classrooms with higher average achievement levels at T1. Students' gains in science vocabulary were further predicted by the share of DLLs and the average SES in the classroom. Thus, above and beyond the role of the average achievement

level in the classroom, students developed their science vocabulary more quickly when grouped together with students from high-SES families and more slowly when they were taught in classrooms with a relatively high share of DLLs (Model 2).

By and large, these findings are mirrored by the results on Topic 2. With regard to students' science content knowledge, prior knowledge was positively related to student achievement at T3. Yet, with 4% of explained variance, the effect of prior knowledge was smaller for "evaporation and condensation" than for "floating and sinking" ($R^2 = 0.11$), which might reflect the test difficulty of the measure on "evaporation and condensation." In line with the findings for "floating and sinking," neither teachers' instructional support nor student composition at the classroom level significantly contributed to students' learning gains.

We additionally investigated the role of concept development, quality of feedback, and language modeling during the teaching unit on "evaporation and condensation" as predictors of students' posttest achievement on science vocabulary, general academic vocabulary, and comprehension of connectives (Models 6–8). Level 1 predictors as well as their classroom aggregates were the same as in Models 3 through 4 (i.e., prior knowledge and sociodemographic characteristics assessed at T1). The observed regression coefficients were thus identical (Level 1) or very similar (Level 2) to those from Models 2 through 4. In line with the findings for Topic 1 but contradictory to our hypotheses, concept development, quality of feedback, and language modeling were not associated with students' posttest achievement on the academic language measures.

Across all eight models, the amount of explained variance varied from $R^2 = 0.07$ for science content knowledge on Topic 2 to $R^2 = 0.26$ for comprehension of connectives at Level 1 and from $R^2 = 0.35$ for science content knowledge on Topic 1 to $R^2 = 0.91$ for science vocabulary at Level 2. Thus, overall, the investigated predictors were considerably more adequate for explaining learning outcomes for the language-related outcomes than for the science-related outcomes.

Discussion

The present study investigated the impact of teachers' instructional support on students' gains in science content knowledge and academic language proficiency in elementary school science classes when controlling for important confounding variables at the classroom level. Analyses were based on data from standardized inquiry-based lesson units on "floating and sinking" and "evaporation and condensation," which were consecutively taught in elementary school science classrooms over 6 months. The study adds to prior research on inquiry-based science instruction in three ways: first, by considering not only the mere effects of curriculum-based interventions, but rather individual differences in teaching behavior; second, by investigating effects on domain-specific and language-related learning gains within regular teaching in mainstream classrooms, thereby considering both general academic language and science vocabulary; third, by its systematic analysis of compositional effects.

Results showed a similar pattern for "floating and sinking" and "evaporation and condensation." We observed a strong impact of students' prior knowledge on science content knowledge (for both "floating and sinking" and "evaporation and condensation") and on students' academic language proficiency. Students' sociodemographic background was also significantly related to their learning gains in some of the outcome variables. Specifically, DLLs showed smaller gains in science content knowledge on "floating and sinking" and

general academic vocabulary, and students with higher SES had better posttest achievement on "floating and sinking" and comprehension of connectives. Moreover, we found compositional effects for the language-related measures, indicating that a high achievement level in class benefits students' academic language development. However, contrary to our expectations, no substantial relations emerged between the three CLASS dimensions of instructional support and any of the elementary school students' learning outcomes.

The finding that instructional support is not associated with students' posttest achievement on content-related and language-related measures contradicts theoretical conceptualizations on classroom quality and the importance of teacher-student interactions (Pianta & Hamre, 2009; Pianta et al., 2008; Praetorius et al., 2018). Thus, along the lines of the theory of social learning (Wood et al., 1976) and the scaffolding approach (e.g., Gibbons, 2002), students' conceptual understanding and language development should benefit from teaching behavior that prompts students to activate prior knowledge, to formulate hypotheses, and to evaluate their assumptions in the light of hands-on activities and observations. When it comes to empirical results, however, prior research provided an inconclusive picture regarding the relationship between instructional support and student outcomes. As for domain-specific knowledge in elementary school science, some studies reported positive effects of instructional support on students' content knowledge (Fauth et al., 2019) and emphasized the role of specific scaffolding strategies and feedback (e.g., Hardy et al., 2006; Decristan et al., 2015). At the same time, a growing body of research conducted in ECEC settings did not identify instructional support as an important driver of children's language development (e.g., Guerrero-Rosada et al., 2021; Perlman et al., 2016; Schmerse, 2021). Whereas the present findings are basically in line with prior results in terms of language skills, they partly diverge from previous findings on the relationship between instructional support and domain-specific knowledge.

Several reasons may account for the minor role of instructional support in the present study. First, we implemented a standardized design with ambitious and effective instructional units on "floating and sinking" (Hardy et al., 2006; Decristan et al., 2015) and "evaporation and condensation." In line with prior research pointing to the general effectiveness of inquiry-based instruction for mainstream classrooms and for ELLs (e.g., Estrella et al., 2018), we observed medium ("floating and sinking") to very large ("evaporation and condensation") effect achievement gains in the posttest. Given the effectiveness of the baseline curriculum and participants' high treatment fidelity in covering necessary lesson content, the potential for an additional impact of instructional support might have been limited (for a similar line of argumentation, see Guerrero-Rosada et al., 2021). Second, in line with prior work that reported null effects of instructional support on preschoolers' language development (e.g., Perlman et al., 2016), the overall quality of instructional support was only mediocre in the present study. Moreover, as reflected in the small correlations between the CLASS dimensions across topics, the present sample showed substantial variability in instructional support over time. Possibly, not only higher levels of instructional support are needed for significant effects on students' domain-specific and language-related learning gains (cf. Burchinal et al., 2010). It can further be assumed that the quality of instructional support and language modeling, in particular, needs to be sustained over longer periods to increase students' (academic) language skills (cf. Alvarez et al., 2012).

In comparing the results of the current study and prior research that implemented similar teaching units, the different assessments of instructional support need to be considered. Whereas previous research mainly drew on researcher-developed observation tools, we employed the CLASS, an internationally validated observation instrument for assessing

interaction quality. Treatment effects tend to be larger on researcher-developed assessments, which are more prone to bias (Oxley & de Cat, 2021), than on well-established, standardized instruments, as the latter were not developed for a particular study (Babinski et al., 2018; Estrella et al., 2018; Kalinowski et al., 2020). The psychometric properties of standardized instruments have typically been widely proven and objectivity is particularly high when engaging independent raters who are not involved in developing or evaluating an intervention (Kalinowski et al., 2020). Using the CLASS, thus, enables greater objectivity, validity, comparability across studies, and international compatibility than using a researcher-developed observation tool. As the CLASS domain of instructional support captures important aspects of interaction quality that were also addressed in our PD programs and respective science units (e.g., linking concepts and activities by conducting experiments, having students think aloud, engaging students in frequent conversations, asking open-ended questions), it is reasonably aligned with our intervention. It should be acknowledged, however, that the CLASS does not explicitly focus on the interplay of language and concept development in inquiry-based learning environments. Moreover, specific language-support strategies, such as the active inclusion of students' multilingual resources or the implementation of reading strategies, are not covered by the CLASS dimension of language modeling. Yet, based on teachers' written documentations about their classroom teaching, we can largely rule out that teachers drew on didactical approaches or teaching materials other than those included in the lesson plans.

While instructional support was unrelated to students' learning outcomes at the posttest, we found pronounced peer effects on students' academic language achievement. Thus, students whose classmates had, on average, higher language proficiency concerning general academic vocabulary and comprehension of connectives at the beginning of the school year showed larger learning gains over time. On the one hand, this finding adds to a large body of research that provided evidence for compositional effects for different age groups and domains, including language proficiency (e.g., Becker et al., 2022; Foster et al., 2020; Hanushek et al., 2003; Schmerse, 2021). On the other hand, it extends prior research on inquiry-based science instruction, which typically did not adhere to compositional effects. Notably, these effects occurred for all language-related measures but for none of the measures on science content knowledge. These diverging patterns of results might be due to design features of the present investigation. Both teaching units aimed at developing students' understanding of challenging concepts and teachers followed detailed lesson plans when delivering the topics in their classes. Although the lesson units provided numerous opportunities for language use and development, lessons had a clear focus on science content knowledge and were led by content-related learning goals. Against this background, the finding that no compositional effects emerged for science content knowledge might point to a buffering effect of the standardized instruction.

Limitations of the present study

Several limitations need to be considered in interpreting the findings of the present investigation. First, the internal consistency of both measures on science content knowledge as well as of the measure of science vocabulary was fairly low in both pretest and posttest, indicating a relatively high measurement error. Yet, low reliabilities are, at least to a certain degree, immanent to the constructs being measured and have thus also been reported in prior research that assessed students' science knowledge (August et al., 2009; Kleickmann et al., 2010; Decristan et al., 2015; Fauth et al., 2014). Given the comparatively small number of items with little redundancy among them, low reliabilities are perhaps an inevitable side effect when assessing

domain-specific content knowledge (cf. Stadler et al., 2021). Yet, all three measures captured learning gains from pre- to posttest on the respective elementary school science topics and the related science vocabulary, thus underlining their sensitivity for assessing effects of instruction.

Second, with 27 classrooms and an average cluster size of 17 students, the present sample was rather small for conducting multilevel analyses. Small sample sizes result in limited analysis power, as reflected in the large standard errors, especially at Level 2, in the current study. While it is unlikely that the sample size affected the overall pattern of results, a larger sample size would have allowed for more complex modeling, including the investigation of cross-level interactions (Hox, 1998).

Third, the assessment of instructional support is based on 20-min cycles of only two double lessons. While it is quite common in educational research that assessments of instructional quality and/or specific teaching strategies are based on a very small number of observations per classroom (Bihler et al., 2018; Guerrero-Rosada et al., 2021; Praetorius et al., 2014; Reyes et al., 2012; Slot et al., 2018) or even on single time points (for an overview of studies, see Praetorius et al., 2014), relying on small cutouts of teachers' classroom behavior certainly comes with constraints. As instructional quality has been shown to vary across lessons, particularly so in lessons focusing on different topics (Hill et al., 2012; Praetorius et al., 2014), using combined measures of multiple observations would help improve the reliability of the assessment.

It should, however, be taken into account that the CLASS ratings in the present study refer to the discussion-intensive introductory sequences of the lessons. By choosing the same sequences for all teachers and across topics, we limited potential variability due to, for instance, differences in teaching arrangements (e.g., independent work in silence vs. group work) or different instructional goals (e.g., learning to spell words correctly vs. learning to formulate hypotheses). Moreover, the selected sequences provided ample opportunity for implementing teaching strategies such as activating prior knowledge, relating content to students' everyday lives, or providing adequate linguistic feedback. Compared to other lesson sequences, they, thus, offered great potential for whole-class support in terms of "concept development," "quality of feedback," and "language modeling."

Conclusions and future directions

There is growing awareness that domain-specific teaching should impact students' content knowledge and improve (academic) language proficiency. Therefore, it is of utmost importance for classroom practice and educational research to develop and evaluate learning environments suitable for integrating language promotion and domain-specific teaching and to prepare teachers for implementing them in their classrooms. Implementing inquiry-based elementary school science classes, we observed substantial learning gains in students' conceptual understanding and academic language proficiency. Teachers' instructional support, however, did not play out on students' learning gains, possibly due to insufficient and altering quality. Given the lack of studies investigating the role of individual differences in teaching behavior for fostering student outcomes in inquiry-based science instruction (Francis & Stephens, 2018), the present study serves as a starting point for future research, focusing more in-depth on the effects of selected teaching strategies. Although the combined measures of concept development, quality of feedback, and language modeling that comprise a number of closely related teaching strategies, were not effective in increasing students' learning gains over and above the effects of the curriculum-based intervention, the selected strategies might indeed be beneficial for specific student groups.

Conducting more fine-grained analyses of selected teaching strategies might therefore provide further insights into effective language-supportive science teaching.

Declarations

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

Conflict of interest The authors declare no competing interests.

References

Adams, R. J., Wu, M. L., Macaskill, G., Haldane, S. A., & Sun, X. S. (2015). *ACER ConQuest: Generalised item response modelling software*. (Version 4) [Computer software]. Australian Council for Educational Research.

Adams, R. J., Wu, M. L., Cloney, D., & Wilson, M. R. (2020). *ACER ConQuest 5. Generalised item response modelling software*. (Version 5) [Computer software]. Australian Council for Educational Research.

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B. K., & Pianta, R. C. (2013). Observations of effective teacher-student interactions in secondary school classrooms: Predicting student achievement with the Classroom Assessment Scoring System—Secondary. *School Psychology Review, 42*(1), 76–98. https://doi.org/10.1080/02796015.2013.12087492

Alvarez, L., Catechis, N., Chu, H., Hamburger, L., Herpin, S. A., & Walqui, A. (2012). *Quality teaching for English learners (QTEL) impact study*. WestEd. https://doi.org/10.13140/2.1.3229.6320

Ardasheva, Y., Newcomer, S. N., Firestone, J. B., & Lamb, R. L. (2017). Mediation in the relationship among EL status, vocabulary, and science reading comprehension. *The Journal of Educational Research, 110*(6), 665–674. https://doi.org/10.1080/00220671.2016.1175407

Ardasheva, Y., Newcomer, S. N., Firestone, J. B., & Lamb, R. L. (2019). Contributions of language‑specific and metacognitive skills to science reading comprehension of middle school English learners. *Bilingual Research Journal, 42*(2), 150–163. https://doi.org/10.1080/15235882.2019.1597774

August, D., Branum‑Martin, L., Cardenas‑Hagan, E., & Francis, D. J. (2009). The impact of an instructional intervention on the science and language learning of middle grade English language learners. *Journal of Research on Educational Effectiveness, 2*(4), 345–376. https://doi.org/10.1080/19345740903217623

August, D., Branum‑Martin, L., Cárdenas‑Hagan, E., Francis, D. J., Powell, J., Moore, S., & Haynes, E. F. (2014). Helping ELLs meet the common core state standards for literacy in science: The impact of an instructional intervention focused on academic language. *Journal of Research on Educational Effectiveness, 7*(1), 54–82. https://doi.org/10.1080/19345747.2013.836763

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. (Ed.). (2007). *The language demands of school. Putting academic English to the test*. Yale University Press. https://doi.org/10.1177/0265532211404199.

Becker, M., Kocaj, A., Jansen, M., Dumont, H., & Lüdtke, O. (2022). Class‑average achievement and individual achievement development: Testing achievement composition and peer spillover effects using five German longitudinal studies. *Journal of Educational Psychology, 114*(1), 177–197. https://doi.org/10.1037/edu0000519

Becker‑Mrotzek, M., & Roth, H.‑J. (2017). Sprachliche Bildung – Grundlegende Begriffe und Konzepte [Language support across the curriculum – Basic terms and concepts]. In M. Becker‑Mrotzek & H.‑J. Roth (Eds.), *Sprachliche Bildung – Grundlagen und Handlungsfelder* (pp. 11–36). Waxmann.

Bihler, L. M., Agache, A., Schneller, K., Willard, J. A., & Leyendecker, B. (2018). Expressive morphological skills of dual language learning and monolingual German children: Exploring links to duration of preschool attendance, classroom quality, and classroom composition. *Frontiers in Psychology, 9*, 888. https://doi.org/10.3389/fpsyg.2018.00888

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

Bunch, G. C. (2013). Pedagogical language knowledge: Preparing mainstream teachers for English learners in the new standards era. *Review of Research in Education, 37*, 298–341. https://doi.org/10.3102/0091732X12461772

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low‑income children in pre‑kindergarten programs. *Early Childhood Research Quarterly, 25*(2), 166–176. https://doi.org/10.1016/J.ECRESQ.2009.10.004

van Buuren, S., & Groothuis‑Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3). https://doi.org/10.18637/jss.v045.i03

Cadima, J., Leal, T., & Burchinal, M. (2010). The quality of teacher‑student interactions: Associations with first graders' academic and behavioral outcomes. *Journal of School Psychology, 48*(6), 457–482. https://doi.org/10.1016/j.jsp.2010.09.001

Decristan, J., Hondrich, A. L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., Lühken, A., Adl‑Amini, K., Djakovic, S.‑K., Mannel, S., Naumann, A., & Hardy, I. (2015). Impact of additional guidance in science education on primary students' conceptual understanding. *The Journal of Educational Research, 108*(5), 358–370. https://doi.org/10.1080/00220671.2014.899957

Eckhardt, A. G. (2008). *Sprache als Barriere für den schulischen Erfolg. Potentielle Schwierigkeiten beim Erwerb schulbezogener Sprache für Kinder mit Migrationshintergrund* [Language as a barrier for academic success. Potential difficulties in acquiring academic language competence for students with a migration background]. Waxmann.

Estrella, G., Au, J., Jaeggi, S. M., & Collins, P. (2018). Is inquiry science instruction effective for English language learners? A Meta‑Analytic Review. *AERA Open, 4*(2), 2332858418767402. https://doi.org/10.1177/2332858418767402

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9. https://doi.org/10.25656/01:18135

Fauth, B., Decristan, J., Decker, A., Büttner, G., Klieme, E., Hardy, I., & Kunter, M. (2019). The effects of teacher competence on student outcomes in elementary science education: The mediating role of

teaching quality. *Teaching and Teacher Education, 86*, 102882. https://doi.org/10.1016/j.tate.2019.102882

Fauth, B., Atlay, C., Dumont, H., & Decristan, J. (2021). Does what you get depend on who you are with? Effects of student composition on teaching quality. *Learning and Instruction, 71*, 101355. https://doi.org/10.1016/j.learninstruc.2020.101355

Fischer, L., Rohm, T., Gnambs, T., & Carstensen, C. H. (2016). *Linking the data of the competence tests (NEPS Survey Paper No. 1)*. Leibniz Institute for Educational Trajectories, National Educational Panel Study. https://doi.org/10.13140/RG.2.2.28728.96004

Foster, T. J., Burchinal, M., & Yazejian, N. (2020). The relation between classroom age composition and children's language and behavioral outcomes: Examining peer effects. *Child Development, 91*(6), 2103–2122. https://doi.org/10.1111/cdev.13410

Francis, D., & Stephens, E. (2018). *English learners in STEM subjects: Transforming classrooms, schools, and lives*. The National Academy Press. https://doi.org/10.17226/2518

Frändberg, B., Lincoln, P., & Wallin, A. (2013). Linguistic resources used in grade 8 students' submicro level explanations—Science items from TIMSS 2007. *Research in Science Education, 43*(6), 2387–2406. https://doi.org/10.1007/s11165-013-9363-0

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research, 82*(3), 300–329. https://doi.org/10.3102/0034654312457206

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

Garza, T., Huerta, M., Lara-Alecio, R., Irby, B. J., & Tong, F. (2018). Pedagogical differences during a science and language intervention for English language learners. *The Journal of Educational Research, 111*(4), 487–496. https://doi.org/10.1080/00220671.2017.1302913

Geary, D. C., Nicholas, A., Li, Y., & Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: An eight-year longitudinal study. *Journal of Educational Psychology, 109*(5), 680–693. https://doi.org/10.1037/edu0000159

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

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2018). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development, 90*(3), 985–992. https://doi.org/10.1111/cdev.13128

Grund, S., Lüdtke, O., & Robitzsch, A. (2017). Multiple imputation of missing data for multilevel models. *Organizational Research Methods, 21*(1), 111–149. https://doi.org/10.1177/1094428117703686

Grund, S., Robitzsch, A., & Lüdtke, O. (2021). *mitml: Tools for multiple imputation in multilevel modeling*. (R package version 0.4–1) [Computer software]. Retrieved August 17, 2022, from https://cran.r-project.org/web/packages/mitml/index.html

Guerrero-Rosada, P., Weiland, C., McCormick, M., Hsueh, J., Sachs, J., Snow, C., & Maier, M. (2021). Null relations between CLASS scores and gains in children's language, math, and executive function skills: A replication and extension study. *Early Childhood Research Quarterly, 54*, 1–12. https://doi.org/10.1016/j.ecresq.2020.07.009

Guo, Y., Wang, S., Hall, A. H., Breit-Smith, A., & Busch, J. (2015). The effects of science instruction on young children's vocabulary learning: A research synthesis. *Early Childhood Education Journal, 44*(4), 359–367. https://doi.org/10.1007/s10643-015-0721-6

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics, 18*(5), 527–544. https://doi.org/10.1002/jae.741

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking". *Journal of Educational Psychology, 98*(2), 307–326. https://doi.org/10.1037/0022-0663.98.2.307

Haug, B. S., & Ødegaard, M. (2014). From words to concepts: Focusing on word knowledge when teaching for conceptual understanding within an inquiry-based science setting. *Research in Science Education, 44*(5), 777–800. https://doi.org/10.1007/s11165-014-9402-5

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

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

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*, 103518. https://doi.org/10.1016/j.tate.2021.103518

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher, 41*(2), 56–64. https://doi.org/10.3102/0013189X12437203

Hox, J. J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 147–154). Springer.

Hu, B. Y., Fan, X., Gu, C., & Yang, N. (2016). Applicability of the Classroom Assessment Scoring System in Chinese preschools based on psychometric evidence. *Early Education and Development, 27*(5), 714–734. https://doi.org/10.1080/10409289.2016.1113069

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*, 4–15. https://doi.org/10.1016/j.tate.2019.102971

Kempert, S., Schalk, L., & Saalbach, H. (2019). Sprache als Werkzeug des Lernens: Ein Überblick zu den kommunikativen und kognitiven Funktionen der Sprache und deren Bedeutung für den fachlichen Wissenserwerb [Language as a tool for learning: A review of the communicative and cognitive functions of language and its relevance for academic learning]. *Psychologie in Erziehung und Unterricht, 65*, 1–20. https://doi.org/10.2378/peu2018.art19d

Kleickmann, T., Hardy, I., Möller, K., Pollmeier, J., Tröbst, S., & Beinbrech, C. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter: Theoretische Konzeption und Testkonstruktion [Modeling scientific competence of primary school children: Theoretical background and test construction]. *Zeitschrift für Didaktik der Naturwissenschaften, 16*, 265–283.

Kleickmann, T., Tröbst, S., Jonen, A., Vehmeyer, J., & Möller, K. (2016). The effects of expert scaffolding in elementary science professional development on teachers' beliefs and motivations, instructional practices, and student achievement. *Journal of Educational Psychology, 108*(1), 21–42. https://doi.org/10.1037/edu0000041

Kohl, K., Willard, J. A., Agache, A., Bihler, L.-M., & Leyendecker, B. (2019). Classroom quality, classroom composition, and age at entry: Experiences in early childhood education and care and single and dual language learners' German vocabulary. *AERA Open*, *5*(1). https://doi.org/10.1177/2332858419832513

Kolen, M. J. (2006). Scaling and norming. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 155–186). American Council on Education.

Kuger, S., Kluczniok, K., Kaplan, D., & Rossbach, H.-G. (2016). Stability and patterns of classroom quality in German early childhood education and care. *School Effectiveness and School Improvement, 27*(3), 418–440. https://doi.org/10.1080/09243453.2015.1112815

Lange-Schubert, K., Böschl, F., & Hartinger, A. (2017). Naturwissenschaftliche Methoden aneignen und anwenden. Untersuchungen durchführen und wissenschaftliche Modelle nutzen am Beispiel Aggregatzustände und ihre Übergänge [Acquiring and applying scientific methods. Conducting experiments and using scientific models based on the example of states of aggregation and transition processes]. In H. Giest (Ed.), *Die naturwissenschaftliche Perspektive konkret. Begleitband 4 zum Perspektivrahmen Sachunterricht* (pp. 25–38). Klinkhardt.

Lara-Alecio, R., Tong, F., Irby, B. J., Guerrero, C., Huerta, M., & Fan, Y. (2012). The effect of an instructional intervention on middle school english learners' science and english reading achievement. *Journal of Research in Science Teaching, 49*(8), 987–1011. https://doi.org/10.1002/tea.21031

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681–718. https://doi.org/10.3102/0034654315627366

Lee, O., & Buxton, C. A. (2013). Integrating science and English proficiency for English language learners. *Theory into Practice, 52*(1), 36–42. https://doi.org/10.1080/07351690.2013.743772

Lee, O., Llosa, L., Grapin, S., Haas, A., & Goggins, M. (2019). Science and language integration with English learners: A conceptual framework guiding instructional materials development. *Science Education*, (103), 317–337. https://doi.org/10.1002/sce.21498

Lee, O. (2020). Science and language instructional shifts with second-language learners. *Asia-Pacific Science Education*, 263-284. https://doi.org/10.1163/23641177-BJA10005

Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., & Rolla, A. (2015). Teacher–child interactions in Chile and their associations with prekindergarten outcomes. *Child Development, 86*(3), 781–799. https://doi.org/10.1111/cdev.12342

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

Lucero, A. (2014). Teachers' use of linguistic scaffolding to support the academic language development of first-grade emergent bilingual students. *Journal of Early Childhood Literacy, 14*(4), 534–561. https://doi.org/10.1177/1468798413512848

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition, 32*(2), 265–302. https://doi.org/10.1017/S0272263109990520

Maerten-Rivera, J., Ahn, S., Lanier, K., Diaz, J., & Lee, O. (2016). Effect of a multiyear intervention on science achievement of all students including English language learners. *The Elementary School Journal, 116*(4), 600–624. https://doi.org/10.1086/686250

Masters, G. N., & Wright, B. D. (1997). The Partial Credit Model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 101–121). Springer New York. https://doi.org/10.1007/978-1-4757-2691-6_6

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

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction-what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47*(4), 474–496. https://doi.org/10.1002/tea.20347

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Muthén & Muthén. Retrieved November 25, 2022, from https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

Ødegaard, M., Haug, B., Mork, S. M., & Sørvik, G. O. (2014). Challenges and support when teaching science through an integrated inquiry and literacy approach. *International Journal of Science Education, 36*(18), 2997–3020. https://doi.org/10.1080/09500693.2014.942719

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

OECD. (2020). Who are the teachers? In Education at a glance 2020: OECD indicators. *OECD Publishing*. https://doi.org/10.1787/27f5f9c5-en

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

Oxley, E., & de Cat, C. (2021). A systematic review of language and literacy interventions in children and adolescents with English as an additional language (EAL). *The Language Learning Journal, 49*(3), 265–287. https://doi.org/10.1080/09571736.2019.1597146

Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J.-E. (2010). A validation of the classroom assessment scoring system in Finnish kindergartens. *Early Education and Development, 21*(1), 95–124. https://doi.org/10.1080/10409280902858764

Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the Classroom Assessment Scoring System) in early childhood education and care settings and child outcomes. *PloS One*, *11*(12). https://doi.org/10.1371/journal.pone.0167660

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System: Manual K-3*. Paul H Brookes.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119. https://doi.org/10.3102/0013189x09332374

Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction, 31*, 2–12. https://doi.org/10.1016/j.learninstruc.2013.12.002

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM Mathematics Education, 50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104*, 700–712. https://doi.org/10.1037/a0027268

Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., Klieme, E., & Stanat, P. (2014). Socioeconomic and language minority classroom composition and individual reading achievement: The mediating role of instructional quality. *Learning and Instruction, 32*, 63–72. https://doi.org/10.1016/j.learninstruc.2014.01.007

Rosén, M., Steinmann, I., & Wernersson, I. (2022). Gender differences in achievement. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), *International handbook of comparative large-scale studies in education. Perspectives, methods and findings.* https://doi.org/10.1007/978-3-030-38298-8_46-1

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. *Wiley.* https://doi.org/10.1002/9780470316696

Sabol, T. J., Bohlmann, N. L., & Downer, J. T. (2018). Low-income ethnically diverse children's engagement as a predictor of school readiness above preschool classroom quality. *Child Development, 89*(2), 556–576. https://doi.org/10.1111/cdev.12832

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education, 12*(4), 431–459. https://doi.org/10.1016/S0898-5898(01)00073-0

Schleppegrell, M. J. (2012). Academic language in teaching and learning: Introduction to the Special Issue. *The Elementary School Journal, 112*(3), 409–418. https://doi.org/10.1086/663297

Schmerse, D. (2021). Peer effects on early language development in dual language learners. *Child Development, 92*(5), 2153–2169. https://doi.org/10.1111/cdev.13588

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

Seah, L. H., & Silver, R. E. (2018). Attending to science language demands in multilingual classrooms: A case study. *International Journal of Science Education, 42*(14), 2453–2471. https://doi.org/10.1080/09500693.2018.1504177

Seah, L. H., Clarke, D. J., & Hart, C. E. (2014). Understanding the language demands on science students from an integrated science and language perspective. *International Journal of Science Education, 36*(6), 952–973. https://doi.org/10.1080/09500693.2013.832003

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2021). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist, Online-Vorveröffentlichung.* https://doi.org/10.1080/00461520.2021.1939700

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

Slot, P. L., Bleses, D., Justice, L. M., Markussen-Brown, J., & Højen, A. (2018). Structural and process quality of Danish preschools: Direct and indirect associations with children's growth in language and preliteracy skills. *Early Education and Development, 29*(4), 581–602. https://doi.org/10.1080/10409289.2018.1452494

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

Snow, C. E., & Lawrence, J. F. (2011). *Word Generation in Boston public schools: Natural history of a literacy intervention* (Vol. 3). The Council of the Great City Schools.

Stadler, M., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: A good alpha is not always better. *New Ideas in Psychology*, *60*, 100832. https://doi.org/10.1016/j.newideapsych.2020.100832

Stuck, A., Kammermeyer, G., & Roux, S. (2016). The reliability and structure of the Classroom Assessment Scoring System in German pre-schools. *European Early Childhood Education Research Journal, 24*(6), 873–894. https://doi.org/10.1080/1350293x.2016.1239324

Tong, F., Irby, B. J., Lara-Alecio, R., & Koch, J. (2014). Integrating literacy and science for English language learners: From learning-to-read to reading-to-learn. *The Journal of Educational Research, 107*(5), 410–426. https://doi.org/10.1080/00220671.2013.833072

Tong, F., Irby, B. J., Lara-Alecio, R., Guerrero, C., Tang, S., & Sutton-Jones, K. L. (2018). The impact of professional learning on in-service teachers' pedagogical delivery of literacy-infused science with middle school English learners: A randomised controlled trial study in the US. *Educational Studies, 45*(5), 533–553. https://doi.org/10.1080/03055698.2018.1509776

van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*, 134–150. https://doi.org/10.1016/j.edurev.2010.02.001

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

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

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

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, *10*, 3087. https://doi.org/10.3389/fpsyg.2019.03087

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

Zwiep, S. G., & Straits, W. J. (2013). Inquiry science: The gateway to English language proficiency. *Journal of Science Teacher Education, 24*(8), 1315–1331. https://doi.org/10.1007/s10972-013-9357-9

Birgit Heppt

Current themes of research:

Assessment and development of academic language proficiency. Teacher professional development for language-supportive teaching. Interplay of language-related and content-related skills.

Most relevant publications in the field of Psychology of Education:

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,* 103518. 10.1016/j.tate.2021.103518

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

Heppt, B. & Stanat, P. (2020). Development of academic language comprehension of German monolinguals and dual language learners. *Contemporary Educational Psychology, 62*, 1-15. 10.1016/j.cedpsych.2020.101868

Heppt, B., Haag, N., Böhme, K., & Stanat, P. (2015). The role of academic-language features for reading comprehension of language-minority students and students from low-SES families. *Reading Research Quarterly, 50*(1), 61–82. 10.1002/rrq.83

Haag, N., Heppt, B., Stanat, P., Kuhl, P., & Pant, H. A. (2013). Second language learners' performance in mathematics: Disentangling the effects of academic language features. *Learning and Instruction, 28*, 24-34. 10.1016/j.learninstruc.2013.04.001

Sofie Henschel

Current themes of research:

Transfer- and implementation research on language and reading support programs. Prerequisites and effects of learning-related emotions.

Most relevant publications in the field of Psychology of Education:

Henschel, S. (2021). Antecedents of science anxiety in elementary school. *The Journal of Educational Research, 1-15*. 10.1080/00220671.2021.1922989

Henschel, S., Rjosk, C., Holtmann, M. & Stanat, P. (2019). Merkmale der Unterrichtsqualität im Fach Mathematik [Features of teaching quality in mathematics]. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 355-383). Waxmann.

Henschel, S. & Roick, T. (2018). The multidimensional structure of math anxiety revisited. Incorporating psychological dimensions and setting factors. *European Journal of Psychological Assessment,* 1-13. 10.1027/1015-5759/a000477

Henschel, S., & Roick, T. (2017). Relationships of mathematics performance, control and value beliefs with cognitive and affective math anxiety. *Learning and Individual Differences, 55,* 97-107. 10.1016/j.lindif.2017.03.009

Henschel, S., Meier, C., & Roick, T. (2016). Effects of two types of task instructions on literary text comprehension and motivational and affective factors. *Learning and Instruction, 44,* 11-21. 10.1016/j.learninstruc.2016.02.005

Ilonca Hardy

Current themes of research:

Adaptive teaching, elementary science education, multilingual education, professional development in inclusion.

Most relevant publications in the field of Psychology of Education:

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,* 103518. 10.1016/j.tate.2021.103518

Hardy, I., Stephan-Gramberg, S. & Jurecka, A. (2021). The use of scaffolding to promote preschool children's competencies of evidence-based reasoning. *Unterrichtswissenschaft,* 49**,** 91–115 (2021). 10.1007/s42010-020-00094-4

Hardy, I., Saalbach, H., Leuchter, M., & Schalk, L. (2020). Preschoolers' induction of the concept of material kind to make predictions: The effects of comparison and linguistic labels. *Frontiers in Psychology, 11:531503,* 1-14. 10.3389/fpsyg.2020.531503

Meschede, N. & Hardy, I. (2020). Selbstwirksamkeitserwartungen von Lehramtsstudierenden zum adaptiven Unterrichten in heterogenen Lerngruppen [Preservice teachers' self-efficacy of adaptive teaching in heterogeneous classrooms]. *Zeitschrift für Erziehungswissenschaft,* 23, 565–589. 10.1007/s11618-020-00949-7

Hardy, I., Decristan, J. & Klieme, E. (2019). Adaptive teaching in research on learning and instruction. *Journal of Educational Research Online*, *11 (2),* 169–191.

Katrin Gabler

Current themes of research:

Language development in school children and development of academic language proficiency. Teacher professional development for language-supportive teaching.

Most relevant publications in the field of Psychology of Education:

Gabler, K., Heppt, B., Henschel, S., Hardy, I., Sontag, C., Mannel, S., Hettmannsperger-Lippolt, R. & Stanat, P. (2020). *Fachintegrierte Sprachbildung in der Grundschule. Überblick und Beispiele aus dem Sachunterricht [Subject-integrated language support in elementary school: Overview and examples from science classrooms].* Berlin: Humboldt-Universität zu Berlin. 10.5159/IQB_ProSach_Handreichung_Lehrkraefte_v1

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, C. Geyer, A. Ropeter, H. Wagner, S. Weber & M. Hasselhorn (Hrsg.), *Sprach- und Schriftsprachförderung wirksam gestalten: Innovative Konzepte und Forschungsimpulse* (S. 59–83). Kohlhammer.

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,* 103518. 10.1016/j.tate.2021.103518