



Cite this: DOI: 10.1039/d6rp00019c

Exploring the impact of socio-scientific issues and scaffolding on student engagement and performance: implications for education for sustainable development

Anna Katharina Liskes * and Helena van Vorst 

This study investigates the impact of the methods of socio-scientific issues (SSI) and scaffolding on students' engagement and performance in chemistry education, with a focus on Education for Sustainable Development (ESD). The research explores in a quasi-experimental intervention with a 2 × 2 pre-/post-design whether the integration of SSI, scaffolding, or a combination of both enhances students' learning outcomes in the context of selected Sustainable Development Goals (SDGs). This is being explored because it is still unclear to what extent ESD learning materials can be effectively implemented in everyday school life and what influence they have on both affective and cognitive learning factors. A total of 662 students from 35 classes in North Rhine-Westphalia, Germany, participated in a 6 week intervention. Using linear mixed effect models, the SSI and scaffolding groups showed significantly higher performance gains compared to the control group, while the students in the combination group did not profit more from the learning material. Engagement was examined both as an overall construct and across five subdimensions (cognitive, emotional, behavioural, agentic, and value-related). Analyses of engagement subscales revealed divergent developmental patterns. Cognitive and agentic engagement showed small but significant increases over time, with gains in cognitive engagement being positively related to performance gains. In contrast, emotional and behavioural engagement declined significantly across measurement occasions, while value-related engagement remained unchanged. These results contribute to the growing body of research on ESD in chemistry education by highlighting the potential of SSI and scaffolding to enhance affective and cognitive learning outcomes. This study underlines the need for further research to optimize the combination of differentiation approaches and explores their potential for fostering long-term performance and engagement.

Received 9th January 2026,
Accepted 18th May 2026

DOI: 10.1039/d6rp00019c

rsc.li/cerp

Introduction

Education for Sustainable Development (ESD) as a concept has been incorporated into the UNESCO education framework as part of the Sustainable Development Goals (SDGs) for school curricula (UNESCO, 2017). Since its introduction in 1992, ESD has received continuous support and is now entering its second phase, aiming for systematic integration across all areas of students' lives, particularly through the 17 SDGs formulated in 2015. Chemistry is particularly well-suited to foster ESD due to its direct relevance to many of the 17 global SDGs. It provides systematic insights into complex causes and underlying factors and offers the potential to transfer the micro-level of chemistry to direct application examples. For instance, SDG 12 – Responsible Consumption and Production – can be effectively addressed in

chemistry education through the topic of plastics. Students can investigate the chemical structure and synthesis of common polymers, examine the environmental impact of plastic waste, and explore strategies for recycling and developing biodegradable alternatives. Experiments on polymer degradation, discussions on life cycle analysis, and critical reflection on consumer habits help students connect chemical knowledge with pressing sustainability issues, such as plastic overconsumption and pollution.

One approach to implement ESD in chemistry education is socio-scientific issues (SSI), which link learners' real-life to subject knowledge (Klosterman and Sadler, 2010; Hofstein *et al.*, 2011) and illustrate the societal relevance of chemistry more clearly, especially in the context of real-world challenges (Stuckey *et al.*, 2013). Previous studies reveal the potential of SSI to increase students' interest in science, which is also a direct aim of the SDGs (Klaver *et al.*, 2023). This seems particularly relevant in view of the fact that chemistry is one of the least popular subjects among students (Hofstein *et al.*, 2011; Avargil *et al.*, 2020).

Didactics of Chemistry, University of Duisburg-Essen, Essen, North Rhine-Westphalia 45127, Germany. E-mail: anna.liskes@uni-due.de



However, when implementing SSI in the classroom, teachers are not only faced with the challenge of creating new learning materials, but also of creating materials that are appealing and supportive for everyone, regardless of their learning background. Scaffolding methods have been established to achieve this goal. Rooted in educational psychology, this approach aims to provide structured support that gradually decreases as learners gain competence (Belland *et al.*, 2022). Through this process, learners develop a systematic understanding that enables them to solve problems, complete tasks, and acquire new skills independently (Wood *et al.*, 1976). For such support to be effective, it must be both adaptive and contingent (van de Pol *et al.*, 2010). This can be achieved particularly through level-adaptive scaffolding, in which tasks and learning materials are differentiated (Belland *et al.*, 2013). In addition to cognitive requirements, scaffolding can also address various affective students' factors and can be used, for example, to increase the students' engagement in the classroom (Azevedo *et al.*, 2004; Acosta-Gonzaga and Ramirez-Arellano, 2022). However, such approaches are still rare and have been the subject of little research. In light of the educational goals outlined in UNESCO's Global Education 2030 Agenda, there is an urgent need to develop clear guidelines, evidence-based recommendations, and practical implementation strategies that make ESD accessible for all students. Scaffolding can help to achieve these requirements.

Based on this background, it is the central aim of this study to investigate the effects of two differentiated approaches, addressing students' performance and their engagement in learning about sustainability. Therefore, a digital learning environment on the topics of plastic and the greenhouse effect was developed for chemistry lessons complying with the guidelines of ESD. Additionally, scaffolding based on students' cognitive prerequisites was implemented by conceptual and strategic scaffolds. In order to accommodate the diverse students' interests, the learning material was embedded in different SSI, among which students could choose during their learning process. This results in two different methods of differentiating the learning material: on the one hand, differentiation based on students' cognitive prerequisites is achieved through integrated scaffolds; on the other hand, differentiation according to students' affective prerequisites is implemented by embedding the learning material in different SSI.

To address this gap, this study investigates whether the implementation of SSI and scaffolding influences students' engagement and performance in chemistry education. Furthermore, it examines whether the combination of both strategies produces interaction effects beyond their individual contributions.

Theoretical frame

Education for sustainable development (ESD)

Global crises, such as pandemics and climate change, are compelling us to reconsider our ways of living and to develop long-term strategies to address these challenges effectively

(Burmeister *et al.*, 2012; OECD, 2020; IPCC, 2022). Hence, ESD has been incorporated into the UNESCO education framework as part of the Sustainable Development Goals (SDGs) for school curricula (UNESCO, 2017). Since its introduction in 1992, ESD has received continuous support and is now entering its second phase, aiming for systematic integration across all areas of students' lives, particularly through the 17 SDGs formulated in 2015 (UNESCO, 2017). The education sector, in particular, is called upon to create learning opportunities that enable all students to engage with these issues in a critical and informed manner. This also entails the active implementation of sustainable changes within the school system. In the context of the 2030 Agenda for Sustainable Development, education is outlined as a stand-alone goal (No. 4) (UNESCO, 2017). While it is an objective in its own right, it also interacts with and supports the achievement of other SDGs by providing a foundational basis (Tahtinen and United Nations Department of Economic and Social Affairs [UN-DESA], 2020). In particular, the UNESCO guidelines offer valuable support and orientation for educators, enabling them to select and adapt appropriate learning contexts for ESD implementation (UNESCO, 2017). ESD identifies explicit learning objectives that can be explored within school curricula, using the SDGs as a guiding framework for fostering sustainable thinking and action (United Nations Department of Economic and Social Affairs, 2025). The objectives are formulated in the cognitive, socio-emotional and behavioural domains to take all dimensions of learning into account. Explicit suggestions and examples of topics are given for each SDG. In contrast to traditional school settings, SDG-related topics are not structured according to age groups or conventional subject areas, but are instead organized around the 17 overarching goals, which promote cross-thematic and cross-curricular approaches to learning (UNESCO, 2017). Nevertheless, many of the topics addressed by the SDGs are also reflected in existing curricula. Integrating ESD in curricula and textbooks is a central goal to implement learning for the SDGs like with the Whole School Approach (WSA) (Mathie, 2024). Especially in light of a rapidly changing world, teachers require clearly developed and empirically evaluated learning environments that help to implement ESD in the classroom and reduce their workload (Waltner *et al.*, 2020; Mulvik *et al.*, 2021). This is precisely by generating insights and tools that are not only academically relevant but also directly applicable and beneficial for teachers in their everyday professional work.

Heterogeneity and differentiated instruction (DI)

It has long been recognized that school classrooms are characterized by a high degree of heterogeneity. Furthermore, studies have shown that not all learners benefit from the same teaching aims, materials and methods (Pozas *et al.*, 2020). A widely used approach to addressing heterogeneity within the classroom is the concept of Differentiated Instruction (DI). DI is a teaching method, in which a teacher takes the abilities or interests of smaller groups of students into account and provides a corresponding learning offer (Tomlinson *et al.*, 2003).



The common practice is to divide learners into groups within the classroom and to adapt learning material in terms of difficulty in order to address students with different abilities (Letzel-Alt and Pozas, 2023). These differences can relate to a wide range of factors, such as prior knowledge, learning pace, reasoning and problem-solving skills, language proficiency, or metacognitive abilities (e.g., Pozas and Schneider, 2019). They may also include affective dimensions such as students' motivation, interest, or engagement in dealing with scientific content (e.g., Potvin and Hasni, 2014). Previous research results show small positive effects for this approach (Hattie, 2009; Sun and Xiao, 2021). Although Hattie (2009) reports rather small effect sizes for ability grouping ($d \approx 0.1$ – 0.2), this does not imply that the practice is inherently ineffective or misguided. Its impact strongly depends on the way it is implemented and on the specific educational context. In almost every classroom, there are substantial differences in students' achievement levels, learning pace, motivation, and language proficiency (Deunk *et al.*, 2018; Smale-Jacobse *et al.*, 2019). A completely uniform approach to teaching would inevitably lead some learners to feel overchallenged while others remain unchallenged. Differentiation is therefore essential to prevent both overload and boredom and to provide all learners with appropriate opportunities for individual learning progress (Bondie *et al.*, 2019). However, many curricula, textbooks, and assessment formats are still designed for relatively homogeneous learner groups (Saleh *et al.*, 2005; Hardy *et al.*, 2019; Langelaan *et al.*, 2024).

Given these challenges, teachers require instructional approaches that allow them to address diverse learning needs without creating excessive workload or stigmatizing students (Gibbs, 2023). Effective strategies must provide a balance between individualization and manageability, enabling teachers to support all learners while maintaining coherent classroom structures (Tomlinson *et al.*, 2003). One promising framework that meets these requirements is the concept of *scaffolding*. It offers a way to operationalize differentiation through structured, adaptive support that adjusts to students' evolving levels of understanding (Belland *et al.*, 2013). Scaffolding provides structured support that helps learners manage complex tasks while gradually fostering increasing autonomy and competence (van de Pol *et al.*, 2010).

Scaffolding

The approach of scaffolding is defined in a relatively uniform manner, drawing on the foundational work of Wood *et al.* (1976). Scaffolding refers to the support provided to students as they solve problems, complete tasks, and develop new competencies. More specifically, scaffolding involves providing adaptive and contingent support tailored to students' current level of understanding and adjusted as learning progresses (Hammond and Gibbons, 2005; Belland *et al.*, 2013). Different types of scaffolding are distinguished in the literature in order to meet different students' needs.

Belland *et al.* (2017) conducted a comprehensive meta-analysis on computer-based scaffolding in Science, Technology, Engineering and Mathematics (STEM) education. Computer-based

scaffolding refers to digital forms of support that assist learners in solving complex, open-ended problems in order to foster higher-order cognitive skills (Belland *et al.*, 2017). The study synthesized findings from 144 experimental investigations comprising a total of 333 individual effect sizes. Across these studies, learners received scaffolded support while engaging with complex, ill-structured problem-solving tasks in instructional settings. This meta-analysis therefore examined the overall effectiveness of digital scaffolding as an instructional strategy across different age groups from elementary to adult education and a wide range of learning contexts. Overall, computer-based scaffolding showed a significant positive impact on students' cognitive learning outcomes. The aggregated effect size was $\bar{g} = 0.46$, corresponding to a medium-to-large effect (Cohen, 2013). This positive influence was consistently observed across different instructional contexts, types of scaffolds, and levels of student performance. A particular focus of the meta-analysis was on the natural science subjects, biology, chemistry, and physics. Among the included studies, 208 effect sizes were related to science instruction. The mean effect size of scaffolding within these science-related contexts was $\bar{g} = 0.42$ (95% CI [0.36, 0.48]). For science education in particular, this indicates that computer-based scaffolding yields similarly positive effects on students' learning outcomes. Belland *et al.* (2017) further revealed positive effects of scaffolding according to the types of cognitive learning outcomes: (1) conceptual and factual knowledge, (2) understanding of principles, and (3) application and problem solving.

In addition, the effectiveness of scaffolding was examined with respect to the nature of the learning task and the instructional context. The studies included in the analysis encompassed a range of learning environments, such as problem-based learning, inquiry-based learning, and general problem-solving tasks (Belland *et al.*, 2017).

Accordingly, scaffolding can be implemented in a wide variety of ways and is often regarded as a complement to existing learning materials. Scaffolds may follow a systematic structure, in which the provided supports are organized according to predefined stages. This enables learners to access assistance step by step and to benefit from different forms of support (Belland *et al.*, 2017). This means that tasks and learning materials are differentiated according to varying levels of difficulty. Additional studies also have demonstrated significant positive learning gains associated with the use of scaffolds, which supports the results so far (van de Pol *et al.*, 2015; Kim *et al.*, 2018; Faber *et al.*, 2024). For example, Broman *et al.* (2018) examined how model-based scaffolds can support upper-secondary students when solving open-ended, context-based chemistry problems. Results show that students without scaffolds tended to remain at low levels of conceptual complexity in their solutions, whereas scaffolded tasks helped many of them reach higher levels of reasoning. The authors conclude that well-designed scaffolds can effectively promote deeper engagement and more advanced chemical thinking in context-based learning environments.

The literature distinguishes various types of scaffolds based on their function and form. Commonly, four major categories



are identified (*e.g.*, Quintana *et al.*, 2004; Azevedo and Hadwin, 2005; Belland *et al.*, 2013):

- conceptual scaffolds, which support understanding of subject matter,
- strategic scaffolds, which guide problem-solving approaches,
- procedural scaffolds, which assist in the use of tools or resources,
- metacognitive scaffolds, which promote reflection and self-regulation.

Empirical evidence indicates no statistically significant differences in learning outcomes among the different types of scaffolds ($p > 0.05$) (Belland *et al.*, 2017). Most approaches to scaffolding do not explicitly distinguish between students who may benefit differently from the provided support. However, previous research has already indicated that such scaffolding holds substantial potential for improving learning outcomes like on students' learning and self-regulation (Belland *et al.*, 2022; Shao *et al.*, 2023).

To maintain the interest and motivation of high-achieving students, it is essential to provide appropriately challenging and enriching learning opportunities. However, inherently focuses on learners in need of support and thus reaches its limits when it comes to addressing the educational needs of high-performing students, who require less assistance but more intellectual stimulation (van de Pol *et al.*, 2010). Expanding the concept of scaffolding beyond its conventional remedial function therefore becomes necessary. Such an expanded perspective should acknowledge that high-achieving learners also benefit from adaptive support, yet not in the form of simplified guidance, but rather through further-thinking scaffolds that open up opportunities for deeper conceptual engagement, advanced reasoning, and autonomous exploration (Acosta-Gonzaga and Ramirez-Arellano, 2022; Belland *et al.*, 2022). In principle, scaffolding offers a flexible framework for this, as it enables adaptive adjustments to learners' varying competence profiles (Chernikova *et al.*, 2025). Nevertheless, empirical research on how scaffolding can be intentionally designed to foster extended engagement and conceptual exploration among advanced learners remains limited. Further studies are needed to investigate how scaffolded extensions, challenge-enhancing prompts, or advanced inquiry elements may encourage these students to work more intensively with scientific ideas and learning materials.

Affective learning outcomes related to scaffolding, such as students' motivation, interest, or changes in attitudes, have received comparatively little attention in the existing research. Belland *et al.* (2017) also emphasized this gap. Their meta-analysis primarily focusses on cognitive performance measures, while affective outcomes were rarely assessed. Only a few studies measure affective variables, mostly motivation (Tuckman, 2007; Rienties *et al.*, 2012; Belland *et al.*, 2013). However, due to the limited number of positive effects of scaffolding on motivation, no generalizable conclusion can be drawn from these findings. In a study by Acosta-Gonzaga and Ramirez-Arellano (2022), the effects of metacognitive and conceptual scaffolding on students' motivation, engagement, and

learning achievement were examined. The research followed a quantitative design using structural equation modelling (SEM). In this study, scaffolding referred to instructional support provided by the teacher, who frames the students' learning process and gradually guides them toward independent learning, rather than the digital or computer-based scaffolding that has been the focus of much prior research. Teacher-provided scaffolding showed a significant positive effect on students' emotional and behavioural engagement. However, no direct influence of scaffolding on learning motivation was found. Instead, motivation acted as a precursor for all forms of engagement. So, motivated students employed more metacognitive strategies, such as self-regulation and reflection, which in turn enhanced their cognitive engagement. Ultimately, students' learning achievement benefited from this increased level of engagement (Acosta-Gonzaga and Ramirez-Arellano, 2022).

Despite the growing body of research on the cognitive effects of scaffolding and a few studies about affective effects, there is also a striking lack of intervention studies that explicitly aim to influence affective factors, although they play a crucial role in determining the depth and sustainability of learning (Acosta-Gonzaga and Ramirez-Arellano, 2022).

Since one of the main aims of this research is to enhance students' engagement by intentionally addressing affective factors, it is essential to select an instructional approach that effectively supports this goal. Previous studies on scaffolding have predominantly employed conceptual, strategic, or metacognitive scaffolds, which are primarily designed to support cognitive processes such as problem solving, reasoning, and self-regulation. Consequently, scaffolding alone may not provide a sufficiently robust framework for systematically and deliberately fostering students' affective characteristics. To address these dimensions more directly, it is reasonable to consider instructional approaches that have been shown to positively impact students' interest, motivation, and engagement. Research indicates that context-based learning through socio-scientific issues (SSI) is particularly effective in this regard, as they connect scientific content to learners' everyday experiences, personal values, and societally relevant questions. These approaches therefore offer promising results for strengthening affective engagement and promoting meaningful science learning. Accordingly, the SSI approach is implemented in this study to complement scaffolding in order to also address the affective dimension in students' learning.

Socio-scientific issues (SSI)

Socio-scientific issues (SSI) are defined as socially relevant, often controversial contexts in science education that integrate scientific concepts with ethical, societal, and sustainability concerns (Sadler, 2011). They not only foster scientific literacy, but also promote critical thinking, social dialogue, and the development of informed and responsible decision-making skills among students (Sadler and Zeidler, 2005). To reach these goals, controversial problems drawn from students' real-life experiences are introduced, encouraging active engagement with scientific content within real-world contexts (Feierabend and Eilks, 2010).



Several models for SSI-based learning materials have been developed, which Sadler *et al.* (2016) summarized under three central principles: (1) the learning material should centre around a complex, socially relevant issue; (2) it should promote engagement in higher-order thinking processes; and (3) both the scientific and social dimensions of the issue should be made explicit. This approach aims to make scientific knowledge personally relevant and socially accepted.

In recent years, research on SSI has increased. A growing number of systematic literature reviews and meta-analyses have been conducted, each focusing on different aspects of SSI. Key themes include decision-making, scientific literacy, and the conceptual definitions of SSI, along with how SSI can be implemented in classroom settings (Tekin *et al.*, 2016; Çalık and Wiyarsi, 2021; Badeo and Duque, 2022; Zulyusri *et al.*, 2022). The following section provides a more detailed presentation of these results.

Generally, the meta-analyses emphasise the potential of SSI for increasing engagement, interest and motivation (Bulte *et al.*, 2006; Parchmann *et al.*, 2006; Romine and Sadler, 2016). Only some of the meta-analyses have shown some first insights into students' engagement with political and societal issues (Schulz *et al.*, 2018; Klaver *et al.*, 2023). Reported effect sizes converge on $\bar{g} \approx 1.1$, *i.e.* learners exposed to SSI progress by one full standard deviation over students taught with conventional approaches (Badeo and Duque, 2022). Badeo and Duque (2022) examined twelve experimental and global SSI interventions in their meta-analysis and found positive effects on students' scientific literacy ($g = 1.08$) or content knowledge ($g = 1.15$), competence ($g = 0.89$), decision-making ($g = 1.14$), and reasoning ($g = 0.81$). The systematic review from Çalık and Wiyarsi (2021) is chemistry-focused and includes 65 papers (2008–2020). The results illustrate that most SSI units foreground immediate, real-life connections, a focus that likely explains their strong motivational impact. The review also highlights those authentic scenarios, such as everyday chemicals, fossil-fuel dilemmas, or microplastic pollution, that interweave multiple relevance components, thereby further enhancing students' perception of significance. Effects are not confined to content recall. They extend to decision-making, argumentative reasoning, and higher-order thinking, demonstrating that SSI pedagogy simultaneously strengthens conceptual and epistemic dimensions of chemistry learning. The strongest effects occur in lower secondary chemistry, where motivation for chemistry typically dips, indicating that SSI can act as a retention level (Çalık and Wiyarsi, 2021).

However, research findings indicate further potential in distinguishing between specific SSI as contexts for science learning, as different learners appear to benefit differently from particular contexts (van Vorst *et al.*, 2015; Habig *et al.*, 2018; Güth and van Vorst, 2023), depending on their prior knowledge and individual interest in chemistry (Güth and van Vorst, 2024). While students with low interest and low performance in chemistry particularly often choose *everyday* contexts (van Vorst and Aydogmus, 2021), students with a comparably higher interest and performance tend to select *uncommon* contexts.

Learners with a high level of interest and high performance in chemistry prefer to choose *laboratory* contexts (Güth and van Vorst, 2024), which address a chemical issue within the subject itself.

However, dealing with SSI in the classroom poses particular challenges due to its inherent complexity, ambiguity, and the need to integrate scientific knowledge with ethical, social, and personal perspectives. These characteristics place high cognitive and emotional demands on students, especially those with lower prior knowledge or less developed reasoning skills. As a result, students often require structured instructional support to engage productively with SSI and to develop informed positions through argumentation and critical reflection. For these reasons, combining SSI and scaffolding may offer a particularly promising approach.

The role of engagement in SSI-based learning

As SSI-based learning environments require students not only to understand scientific content but also to evaluate complex, real-world problems, students' personal involvement, critical thinking, and autonomous decision-making are necessary to solve SSI-based problems appropriately (Appleton *et al.*, 2008). For this reason, it is essential for the successful implementation of SSI in the classroom to promote students' engagement in addition to scientific content knowledge. Engagement plays a central role in both educational and psychological research, as it serves as a key mediator between motivational processes and actual learning outcomes (Azevedo, 2015). It reflects the extent to which students invest cognitive, emotional, and behavioural resources in the learning process. Accordingly, the concept of engagement in educational and psychological research is generally understood as a multidimensional construct encompassing behavioural, emotional, and cognitive components (Finn and Zimmer, 2012; Azevedo, 2015; Sinatra *et al.*, 2015). Behavioural engagement refers to students' participation, effort, and persistence in academic activities. Emotional engagement captures affective reactions such as interest, enjoyment, or a sense of belonging that accompany learning processes. Cognitive engagement involves the investment of mental effort in understanding complex ideas, employing deep learning strategies, and engaging in self-regulated thinking. Together, these dimensions describe the quality and intensity of students' active involvement in learning and are considered essential indicators of meaningful educational participation. From a psychological perspective, engagement represents the visible manifestation of motivation (Azevedo, 2015; Ryan and Deci, 2020). In classroom contexts, engaged students are more likely to demonstrate curiosity, self-regulation, and resilience when faced with challenging content. In contrast, low engagement often leads to surface-level learning and disengagement from science as a discipline (Inkinen *et al.*, 2020; Schnitzler *et al.*, 2021). For this reason, fostering engagement is considered essential for promoting both cognitive growth and positive affective development in science education (Fredricks *et al.*, 2004; Azevedo, 2015).

This perspective allows for a nuanced consideration of the level or degree of student involvement or participation within



the classroom context (Sinatra *et al.*, 2015; Li *et al.*, 2023). However, the concept of engagement used in this study extends beyond the traditional three-dimensional model (Fredricks *et al.*, 2004). In subsequent theoretical developments, additional facets have been incorporated to capture more nuanced forms of students' active participation. Reeve (2013), for instance, introduced the notion of agentic engagement, which refers to students' constructive contributions to the flow of instruction, such as expressing preferences, asking questions, or proposing ideas that influence their own learning environment. This dimension emphasizes learners' proactive role in shaping educational interactions rather than merely responding to them. It also reflects the degree to which students autonomously motivate themselves and actively participate in shaping their learning (Reeve and Shin, 2020; Reeve *et al.*, 2022). Moreover, the concept of engagement, as traditionally conceptualized, exhibits notable parallels to Krapp's (2002) theory of interest, particularly in the attribution of an emotional component to the state of being engaged. Drawing on Krapp's (2002) theory of interest, the present study therefore extends the operationalization of engagement by incorporating the value-related component of interest as an additional dimension (*value-related engagement*). This inclusion is grounded in the assumption that the subjective value ascribed to a learning situation and its content may likewise foster higher levels of engagement in class. Value-related engagement refers to the perceived personal relevance and importance of learning activities (Fechner, 2009). Together, these extensions offer a more comprehensive framework for analysing students' multifaceted involvement in learning processes.

Aims and research question

Scientific topics coming along with ESD are often complex and require substantial background knowledge. To enable students to engage competently with these complex issues, both performance and engagement must be explicitly promoted. Furthermore, students need additional instructional support through differentiated learning approaches. The concept of scaffolding has already shown positive effects on learning outcomes in previous studies (Belland *et al.*, 2017; Shao *et al.*, 2023; Faber *et al.*, 2024). Nevertheless, its implementation, especially in chemistry education, has been rather limited, and only a few large-scale studies have systematically examined its potential impact. The influence of scaffolding on affective parameters has also received little attention so far (Belland *et al.*, 2017). The integration of SSI with different thematic contexts has been shown to positively influence students' engagement (Feierabend and Eilks, 2010; Badeo and Duque, 2022; Klaver *et al.*, 2023). However, a clear research gap remains, as no studies have yet examined this relationship explicitly, particularly regarding whether ESD-related materials based on SSI can achieve such effects. Moreover, the combination of both approaches, SSI and scaffolding, to simultaneously address affective and cognitive aspects of learning has not yet been investigated. Hence, the

present study aims to investigate the effectiveness of scaffolding as an approach to differentiation based on learners' cognitive prerequisites, in comparison to differentiation grounded in affective learner characteristics, implemented using distinct SSI for ESD. In the present study, scaffolding and SSI were conceptualised as distinct instructional design principles. As an additional result, instructional materials that allow teachers to directly implement empirically tested, curriculum-aligned ESD content in their classrooms will be developed and evaluated.

On the one hand, scaffolding will be applied to provide cognitive support, ensuring that both lower- and higher-achieving students can benefit from the learning materials. This allows for the investigation of whether scaffolding effectively enhances students' performance across different performance levels. On the other hand, SSI will be implemented to address affective factors, specifically, students' engagement, with the aim of increasing their proactive attitude and involvement in chemistry learning.

Furthermore, the study will examine whether a combination of both approaches, scaffolding and SSI, can jointly foster cognitive and affective outcomes. Given the considerable heterogeneity among students in both cognitive ability and affective disposition, this investigation will provide valuable insights into how diverse learners can be effectively supported in ESD-oriented chemistry instruction. With the help of a 2×2 study design, the effects of these two methods of differentiation are investigated, resulting in the following research question (RQ):

To what extent do ...

- (a) socio-scientific issues (SSI),
- (b) scaffolding,
- (c) SSI and scaffolding in combination

... influence students' engagement and their performance in chemistry education?

Based on the described state of research, the following hypotheses can be formulated accordingly:

- (1) Engagement is positively influenced by SSI (Bennett and Lubben, 2006; Badeo and Duque, 2022)
- (2) Performance is positively influenced by scaffolding. (*e.g.*, Bennett *et al.*, 2016; Belland *et al.*, 2017)
- (3) Engagement and performance are positively influenced by the combined use of SSI and scaffolding.

Methods

Material development

In light of the need to develop materials that align with ESD, the school curriculum in organic chemistry was recently expanded (MSB KLP, 2019). As a result, new learning materials had to be developed, because no existing resources offered appropriate options for differentiated instruction. The development of the learning materials in this study is based on the concept of Ladders of Learning (LL), which offers a clear, structured, and student-centred framework for guiding individual learning processes (van Vorst, 2018). In its original form, an LL represents both the content and methods of instruction



through a sequence of structured units arranged in a linear progression (Müller *et al.*, 2015). Each content area is divided into thematic sections, referred to as *milestones*, which correspond to the rungs of a ladder. Each milestone consists of several instructional units that typically follow a consistent pattern:

- (1) Introductory – acquisition of new knowledge;
- (2) Reinforcement – independent use and consolidation of knowledge;
- (3) Evaluation – diagnosis of potential knowledge gaps;
- (4) Remedial work – targeted exercises to close gaps;
- (5) Enrichment – transfer and deeper application of acquired knowledge.

Building on the structure from van Vorst (2018), the present study adapts the LLs approach to a digital learning material on tablets that can be accessed quickly and used across diverse classroom settings. Digital implementation enables the integration of multimedia elements, interactive scaffolds, and adaptive feedback systems that support differentiated learning pathways (Maier and Klotz, 2022; Reinhold *et al.*, 2024). Furthermore, the digital LL offers the potential for continuous empirical evaluation and iterative improvement, ensuring that ESD-related content could be implemented efficiently and sustainably in contemporary chemistry education.

The learning material for this study contained three consecutive learning units on the subject of organic chemistry:

Milestone 1: polymers and their properties

Milestone 2: a recycling cycle

Milestone 3: the greenhouse effect

These topics are mandatory for the third year of chemistry instruction according to the current curriculum (MSB KLP, 2019). Beyond this, they also address current and socially significant issues directly linked to the goals of ESD. Specifically, the learning material contributes to SDG 12 – Responsible Consumption and Production, by encouraging students to critically reflect on the life cycle and environmental impact of synthetic materials, and to SDG 13 – Climate Action, by connecting chemical principles with climate-related phenomena such as the greenhouse effect (UNESCO, 2017). In this way, the LL helps students understand how chemistry can contribute to sustainable solutions for global challenges.

The learning material included phases of independent work with digital materials as well as structured partner and small-group activities. In addition to text-based tasks, the materials incorporated integrated hands-on activities related to the respective content like conducting experiments in the lab or creating models as visualisations. Thus, students engaged with the subject matter through a combination of individual, collaborative, and practical work formats within the intervention.

To support students in the learning process, continuous feedback was integrated in the LL. At the beginning of the first milestone, students receive initial feedback based on their results in the prior knowledge test, indicating which level of optional support they might choose. After each reinforcement phase and each evaluation, students receive additional feedback, allowing for a gradual adjustment in the extent to which the scaffolds were used throughout the learning process.

For each task, scaffolds were integrated optionally, as the learning material could also be completed without them. All scaffolds were available for every student and were placed next to the task. Students could click on a scaffold if they wanted to and were allowed to use more than one scaffold. Different scaffold formats were implemented:

Strategic scaffolds were used to provide initial assistance by encompassing techniques such as paraphrasing, planning partial steps or activating prior knowledge. For example, relevant information from previous learning materials was explicitly referenced to help students complete the current task. In addition, the task itself may have been pre-structured, allowing learners to organize information more easily, for instance, by filling in a prepared table or by receiving suggested text elements as guidance.

Conceptual scaffolds that provide supplementary information about the learning content were integrated, like clarifications of key terms, references to relevant formulas, data, or scientific laws. In these scaffolds, students received the necessary information directly. Additionally, the theoretical content was presented in a simplified and condensed form, focusing on the essential concepts, *e.g.* by using fill-in-the-blank exercises or other supportive formats that were meant to reduce cognitive load. Supplementary visual elements, such as models, photos, and illustrations, are also included to support the content.

Support for further thinking was provided to offer challenges for high-achieving students. These scaffolds included additional application scenarios, problem-solving tasks, or opportunities for generalized learning, and, hence, engaged high-performing students more deeply with the topic. Fig. 1 and 2 show an example task that illustrates the three different types of scaffolding. In the actual learning environment, selecting a scaffold revealed only the respective support. For illustrative purposes in this article, all scaffolds are shown simultaneously. In this example task, students are encouraged to engage in further reasoning by developing ideas for a simple experiment that could be used to test the theoretical explanation addressed in the task. This element represents the scaffold for further thinking, as it requires students to apply the acquired concepts to a new situation and design a possible experimental

Milestone 4 Introductory The Greenhouse Effect

c) Now think back to the experiment with the ice cubes from the beginning. Use the greenhouse effect to explain the observations you made there. Then *explain* why tomatoes can be grown in greenhouses even in winter.

Experiment Setup
"Ice Cubes in a Glass in the Sun"

← BACK Overview NEXT →

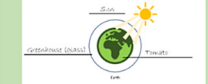
Fig. 1 Example task from the learning material with the three scaffolding types on the right-hand side: support for further thinking (head with brain), strategic (construction cone), and conceptual (question mark).



Milestone 4 Introductory **The Greenhouse Effect**

Collect ideas for an experiment that could help you test your explanation.

Take a look at this tip.



Complete the following sentences using the keywords to answer the question.

The ice cube in the closed beaker melts faster than in the Petri dish because [temperature – glass – higher] _____. As a result, it can [release – less – heat – to the surroundings] _____. Since the greenhouse is also enclosed, [escapes – less – heat – environment] _____, so that [tomato plants – winter – better – grow] _____.

BACK Overview

Fig. 2 Example task from the learning material with the three scaffolding types on the right-hand side: support for further thinking (head with brain), strategic (construction cone), and conceptual (question mark).

approach. The strategic scaffold provides a visual hint referring back to a previous task, prompting students to transfer their knowledge from the earlier thematic context to the greenhouse effect discussed in the current task. In contrast, the conceptual scaffold supports students by providing a short explanatory text in which key terms only need to be integrated into a correct sentence, thereby reducing the complexity of the task and facilitating conceptual understanding. This is an example, so the support methods vary in the tasks and can also take different forms as described, so that the learning material remains varied.

In contrast to the differentiation provided through scaffolding, different SSI were implemented as an overarching thematic layer across each milestone (Table 1). Within every milestone, students could choose between three different thematic SSI in which to work through the learning material.

All SSI were broadly related to the overarching theme of food and nutrition and, in line with the literature (van Vorst *et al.*, 2015; Sadler *et al.*, 2016; Badeo and Duque, 2022) presented earlier, represented three types of contexts: one focused on an everyday-life issue (everyday context), another one addressed a special and socially relevant phenomenon (uncommon context), and a third context focused on a subject-specific scientific application in the lab (laboratory context). It should be noted that, although the curriculum defines overarching topics in organic chemistry (*e.g.*, polymers or the greenhouse effect), it does not prescribe specific instructional contexts. The selection and implementation of contexts are therefore part of the instructional design rather than

Combination group (S+SSI): (+) Scaffolding (+) SSI $n_{S+SSI} = 186$	Scaffolding group (S): (+) Scaffolding (-) SSI $n_S = 184$
SSI group (SSI): (-) Scaffolding (+) SSI $n_{SSI} = 160$	Control group (CG): (-) Scaffolding (-) SSI $n_{CG} = 132$

Fig. 3 Intervention groups of the 2×2 -design in the intervention study.

inherent to the curriculum. In the present study, the same curricular topics were addressed across conditions, while the contextual framing differed systematically.

The selection of these SSI aligns closely with several SDGs, particularly SDG 2 – Zero Hunger, SDG 12 – Responsible Consumption and Production, SDG 13 – Climate Action, and SDG 14 – Life Below Water. These connections reflect the relevance of sustainable food consumption, the use and recycling of plastics in everyday life, and the environmental consequences of plastic pollution in marine and terrestrial ecosystems (UNESCO, 2017). After completing each milestone, students were free to choose a different SSI for the next milestone.

However, it is essential to emphasize that all students engaged with identical content and tasks, while only the thematic framing of the learning tasks varied. To maintain comparability across conditions, a uniform layout was used for every SSI, preventing potential bias in students' choices due to variations in text amount or presentation style.

Study design and instruments

To answer the research question, a quasi-experimental intervention study in a pre-/post-design was conducted in 13 upper secondary schools (German: Gymnasium) across North Rhine-Westphalia, Germany. A total of 20 teachers participated with their chemistry courses ($n = 35$ courses) voluntarily. The courses were randomly assigned to one of four intervention groups in advance, illustrated in Fig. 3. Students in the intervention group S + SSI received both scaffolding and SSI. Students in intervention group S only received scaffolding while students in the intervention group SSI were able to choose between different SSI but without scaffolds. The fourth intervention group served as a control group (CG) in which neither

Table 1 SSI thematic differentiation through three different contexts for each milestone

Milestone	Context		
	Everyday	Uncommon	Laboratory
1	Plastic packaging in our everyday lives	The appearance of garbage islands due to plastic packaging	Use of plastic in the laboratory
2	Recycling food packaging in Germany	Recycling plastic to combat pollution in the world's oceans	Recycling plastics from the laboratory
3	How do your eating habits affect the Earth's climate?	Algae as the food of the future: Can it solve the problem of climate change?	Demonstrating the greenhouse effect in the laboratory



Pre-test	Intervention				Post-test
1 lesson; 45 minutes	3 lessons	3 lessons	4 lessons		1 lesson; 45 minutes
	Milestone 1: polymers and properties	Milestone 2: recycling cycle	Milestone 3: greenhouse effect		
	Intermediate test (5-10 minutes)	Intermediate test (5-10 minutes)	Intermediate test (5-10 minutes)		

Fig. 4 Proceedings of the conducted intervention.

SSI nor scaffolding were available. The control group worked with tasks embedded in the laboratory context. This design decision was made because previous research suggests that this type of context is the least frequently chosen by students when alternatives are available (Güth and van Vorst, 2024). At the same time, laboratory contexts closely resemble the traditional, content-focused approach commonly used in chemistry instruction. Using this context for the control group therefore allows for a baseline condition that reflects typical classroom practice, while minimizing additional motivational influences that might arise from more engaging or personally relevant contexts. In general, all students worked on the same content and tasks, ensuring that potential differences in learning outcomes could be attributed to the experimental conditions rather than variations in learning material. SSI and scaffolding were implemented as independent experimental variables within the 2×2 design.

Fig. 4 summarises the proceedings of the intervention. Before the intervention started, students completed a pretest during their regular chemistry class. This test measured conceptual and factual knowledge relevant to the LL and took about 45 to 60 minutes. As a dependent variable students' content knowledge in organic chemistry was assessed with a multiple-choice single-select test (Celik, 2022; self-developed). The core instrument was previously developed and validated by Celik (2022). However, due to the recent extension of the school curriculum (MSB KLP, 2019), the original instrument did not include items covering newly introduced organic chemistry content. Therefore, additional items were developed in alignment with the updated curricular specifications ($n = 21$). These newly developed items were evaluated in a pilot study prior to the main intervention. Item quality was examined using standard procedures of test construction, including analyses of item difficulty and discrimination (Bond, 2015). Reliability of the pilot instrument was assessed using expected a posteriori (EAP) reliability, which indicated good internal consistency (EAP = 0.88). In addition to psychometric criteria, item selection was guided by content considerations. As the pilot instrument covered a broader range of topics (e.g., alkanes and alkanols), only those items were retained that aligned with the content focus of the main study ($n = 11$). This selection was informed by the empirically validated learning progression from Celik (2022), so that the test now comprises $N = 18$ items. However,

the remaining items were used in the prior knowledge test ($n = 10$). In total, there are $N = 23$ items. The exact distribution can be found in Table 2.

Additionally, students' engagement in chemistry was assessed with an adapted version of the engagement questionnaire by Li *et al.* (2023). The instrument comprises four subscales representing emotional, cognitive, behavioural, and agentic engagement. To extend the theoretical framework, items on value-related valence were included, adapted from Fechner (2009), to capture the perceived relevance of learning content. All items were rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The engagement subscales were examined using confirmatory factor analysis (CFA). First, a CFA was conducted at pre-measurement to test the hypothesised five-factor structure of the engagement questionnaire (cognitive, behavioural, emotional, agentic, and value-related engagement). The model showed acceptable fit to the data, $\chi^2(199) = 1066.85$, $p < 0.001$, CFI = 0.85, TLI = 0.82, RMSEA = 0.081, and SRMR = 0.078. All items loaded significantly on their respective factors, with standardised loadings ranging from 0.30 to 1.16. Subsequently, a longitudinal CFA was estimated to examine the stability of the measurement model across time (pre- and post-measurement). Given the ordinal nature of the indicators, models were estimated using the diagonally weighted least squares (DWLS) estimator. The configural model, specifying the same factor structure across time, showed an acceptable fit to the data, $\chi^2(857) = 2898.51$, CFI = 0.885, TLI = 0.873, RMSEA = 0.072, and SRMR = 0.075. Metric invariance was tested by constraining factor loadings to be equal across time. Model comparisons indicated only negligible changes in fit ($\Delta\text{CFI} = 0.009$, $\Delta\text{RMSEA} = -0.004$), supporting metric invariance. Subsequently, threshold invariance was examined by additionally constraining item thresholds across time. This more restrictive model showed only minor changes in fit ($\Delta\text{CFI} = -0.01$, $\Delta\text{RMSEA} = 0.000$), suggesting that threshold invariance can be assumed. Although the chi-square difference test was statistically significant, $\Delta\chi^2(88) = 430.37$, $p < 0.001$, this result was expected given the large sample size. Therefore, the evaluation of invariance was primarily based on changes in approximate fit indices, which supported the assumption of longitudinal measurement invariance (Meredith, 1993; Chen, 2007; Brown, 2015). Based on these results, measurement invariance was supported, and all items were retained. Consequently, the five-factor structure was maintained for subsequent analyses.

As control variables, students' descriptive and background data were surveyed in the pre-test, including gender, school grade, and course choice in upper secondary education. In addition, students' prior knowledge in chemistry, relevant for understanding the chemical content in the LL, was measured using a multiple-choice single select test (self-developed; Celik, 2022). After the intervention, dependent variables (content knowledge in organic chemistry and engagement) were assessed again, using identical instruments as in the pretest. Additionally, students' satisfaction with the learning environment (adapted from Engeln, 2004) and its usability (adapted



Table 2 Overview of instruments, subscales, and their reliabilities

Time	Instrument	Variable type	Subscale(s)	No. of items	Example item	Reliability (α)
Pre and Post	Performance test (adapted from Celik, 2022; self-developed)	Dependent (cognitive) 4-multiple choice, single select	Organic chemistry knowledge; $n = 7$ (adapted from Celik, 2022); $n = 11$ (self-developed)	18	“What are the main differences between thermoplastics, duroplastics and elastomers?”	Pre = 0.59, post = 0.81
Pre and post	Engagement questionnaire (adapted from Fechner, 2009; Li <i>et al.</i> , 2023)	Dependent (affective) 5-likert scale		22		Pre = 0.90, post = 0.93
			Cognitive	5	“In chemistry class, I try to connect what I’m learning with what I already know.”	
			Emotional	5	“I enjoy learning new things in chemistry class.”	
			Behavioural	5	“In chemistry class, I listen attentively to my teachers and classmates.”	
			Agentic	5	“In chemistry class, I make suggestions on how to solve the problems more effectively.”	
			Value-related	2	“I find that what I learn in chemistry class is very useful.”	
Pre	Demographic data	Control	Gender, grade, course choice	3	“What is your gender?”	—
Pre	Prior knowledge test (adapted from Celik, 2022; self-developed)	Control/covariate 4-multiple choice, single select	General chemistry knowledge; $n = 13$ (adapted from Celik, 2022); $n = 10$ (self-developed)	23	“Which of the following interactions is an intermolecular interaction?”	0.64
Post	Satisfaction with learning environment (Güth and van Vorst, 2024)	Control 5-likert scale		4	“I was satisfied with the topic of my task.”	0.92
Post	Usability scale (Brooke, 1995)	Control 5-likert scale		10	“I think working in the online learning environment is easy.”	0.39

from Brooke, 1995) were measured. Unfortunately, it was not possible to include tracking data on individual scaffold usage as students work on their own digital devices. Table 2 gives an overview of the instruments used in this study. While most scales demonstrated high reliability, the usability scale showed poor internal consistency. Therefore, no further statistical analysis will be conducted for these measures. This low reliability likely resulted from an insufficient distinction between satisfaction with the learning material itself and the technical aspects of the digital tool used to present the learning material.

Procedure and participants

The intervention was conducted during 12 regular chemistry lessons (45 minutes each) and was integrated in regular chemistry courses in school. Students worked on the LL every week during their regular chemistry lessons on their own tablet computers. The teacher primarily assumed a facilitating and organizational role, providing guidance on the overall workflow, answering questions when necessary, and ensuring that all students were at the same progress level at the end of each lesson. A detailed lesson plan was provided to the teacher to support this process. Pre- and post-tests were administered in the presence of a trained research assistant to ensure that all test items were completed individually and anonymously. This research assistant was available for any technical, organisational, and content-related questions of the teachers, also about the whole execution time both by telephone and email.

A total of $N = 1177$ students started with the intervention, while $n = 662$ complete data sets could be used for further analysis.

The remaining students were distributed almost evenly between the intervention groups, $n_{S+SSI} = 186$, $n_{SSI} = 160$, $n_S = 184$, and $n_{CG} = 132$. The students’ age was around 15 years and about 52.6% of the students were female. The school grade was within the expected range, which was between 2 and 3[†] (SD = 0.99).

The relatively long duration of the intervention and the limited time available before the summer holidays meant that some classes were unable to complete all learning units. For this reason, $n = 259$ students were sorted out by class. In some cases, individual students missed either the pre- or post-tests due to illness, school absence, or class changes, which led to their exclusion from the dataset ($n = 215$). Other external factors also played a role in the data loss. The intervention took place during the second half of the school year, when chemistry lessons were typically taught two lessons per week, and in some schools only one lesson per week due to organizational constraints. School vacations, public holidays, excursions, and other school events further reduced the available time, creating additional time pressure for an already extensive intervention. So, $n = 703$ students remain in the data set for the time being.

[†] In Germany, the grading system typically ranges from 1 to 6, where 1.0 represents “very good,” 2.0 “good,” 3.0 “satisfactory,” and 4.0 “sufficient.” Grades of 5.0 indicate “poor” performance, while 6.0 represents “fail.”



Additionally, students who completed the learning material in an unusually short amount of time were excluded, as it was assumed that they had clicked through the digital material without proper interaction or reflection ($n = 41$). As the last exclusion comprised only about 3% of the dataset, it was considered unlikely to introduce systematic bias or meaningfully affect the overall results (Little and Rubin, 2002).

Consequently, most of the missing data can be attributed to external organizational factors or student absences rather than systematic bias. Therefore, the data were assumed to be missing completely at random (MCAR) (Barnes *et al.*, 2008). Based on this assumption, listwise deletion was applied to cases with missing values in order to avoid biasing the dataset. Estimating missing values was not considered appropriate under the MCAR condition.

Data analyses

Given the hierarchical structure of the collected data and the 2×2 factorial design, a linear mixed-effects model (LMM) is an appropriate method for statistical analysis (Tijmstra and Bolsinova, 2023; Molenberghs and Verbeke, 2025). This approach accounts for both fixed and random effects, allowing the model to control for dependencies, in this case, students nested within classes. Since randomization was carried out at the class level and a 2×2 factorial design was implemented, LMMs are particularly well-suited to disentangle treatment effects from classroom-level variance and to increase the precision of estimated effects by modelling intra-class correlations (Lazega and Snijders, 2016). This approach allows for the examination of main and interaction effects between the two factors. This is particularly important when analysing group differences in dependent variables, as it prevents inflation of Type I error rates that can occur when ignoring data dependency. Thus, the use of LMM in this study provides a statistically sound framework to test for differences and interactions in engagement and performance across the experimental groups while appropriately handling the data structure. In addition, linear mixed models are robust to unequal sample sizes across treatment groups, as parameter estimation is based on (restricted) maximum likelihood, allowing all available data to contribute proportionally to the model estimates.

In accordance with common practice in the analysis of linear mixed-effects models, this study adopts the conventional significance level of $\alpha = 0.05$ for hypothesis testing (Luke, 2017). This threshold balances the risk of committing a Type I error (false positive) with the practical need for detecting meaningful effects. However, given the factorial design and multiple comparisons involved in examining both main and interaction effects, there is an increased risk of inflating the overall Type I error rate. To address this issue, an alpha error correction method will be applied to control for multiple testing. Such corrections reduce the likelihood of false positives by adjusting the threshold for statistical significance. Common approaches include the Bonferroni correction or the False Discovery Rate (FDR) procedure (Benjamini and Hochberg, 1995), which

maintain the family-wise error rate or the expected proportion of false discoveries within acceptable limits. Employing these corrections enhances the robustness of the findings by ensuring that significant results are not due to chance alone, thus increasing the reliability and validity of the conclusions drawn from the data.

Findings

To ensure the comparability of the intervention groups, a multivariate analysis of variance (MANOVA) including the control variables was conducted at the beginning of the data analysis. A MANOVA allows the simultaneous examination of group differences across several dependent variables while controlling for potential intercorrelations among them. This approach provides a more comprehensive assessment of overall group comparability than separate univariate analyses (Field, 2024). There are no significant differences in gender, school grade and course choice for upper secondary school between students in the four intervention groups. There are significant differences between the intervention groups in pre-performance. Group S + SSI has a significantly higher test score than the other intervention groups, $F(3, 699) = 4.44, p = 0.004, \eta^2 = 0.02$. Also, in engagement, there are significant differences at the pre-test, $F(3, 658) = 2.63, p = 0.049, \eta^2 = 0.01$. The control group reports about a significantly higher engagement than group S + SSI. Hence, students' engagement and their pre-test results in the performance test have been used as covariates in the following analyses.

Fig. 5 and 6 present the descriptive statistics for the measured variables' performance and engagement across the four intervention groups. The descriptive data show an increase from pre- to post-test for performance for all groups, suggesting general learning progress across conditions. Group SSI shows the highest post-test mean ($M = 10.73, SD = 3.56$), whereas the control group exhibits slightly lower performance gains

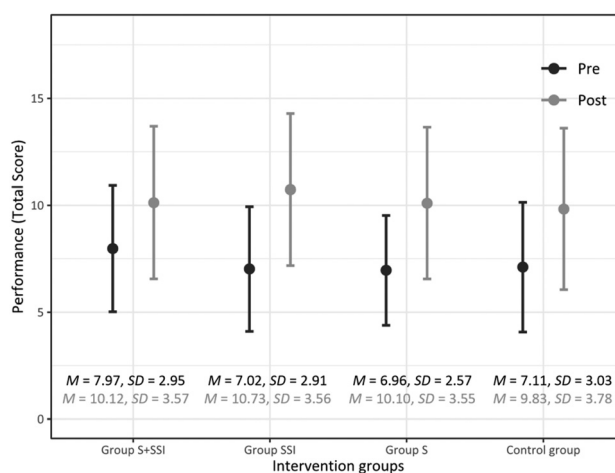


Fig. 5 Descriptive statistics for performance (total score) with means (M) and standard deviation (SD) for all intervention groups pre and post. MT = Measurement, T = Treatment, S = Scaffolding, SSI = Socio-scientific Issues.



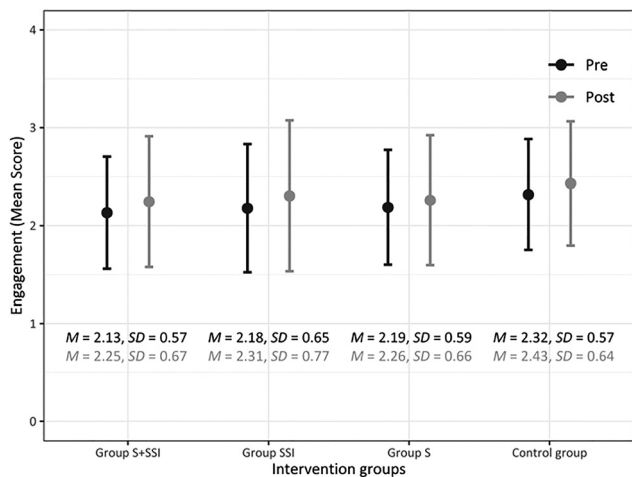


Fig. 6 Descriptive statistics for engagement (mean score) with means (M) and standard deviation (SD) for all intervention groups pre and post. MT = Measurement, T = Treatment, S = Scaffolding, SSI = Socio-scientific Issues.

($M = 9.83$, $SD = 3.78$). In terms of engagement, mean values are relatively stable across groups, with small increases from pre- to post-measurement.

Detailed descriptive statistics for the five engagement subscales (emotional, cognitive, behavioural, agentic, and value-related) are reported in the Appendix (Fig. 7–11). The mean values are relatively stable across groups, with partly small increases from pre- to post-measurement. On average, the mean scores vary somewhat. For example, the mean score for emotional engagement in group S + SSI ($M = 2.56$, $SD = 0.71$) is about one point higher at the pre-measurement than that for agentic engagement ($M = 1.23$, $SD = 0.79$). However, there is no significant change over time in any engagement scale.

Based on these descriptive results, LMMs were calculated to reveal differences between the intervention groups. The LMMs reported here use the control group as the reference category, as this comparison most directly addresses the research question. For the sake of statistical completeness and to verify the robustness of the findings, additional models were estimated in which each intervention condition served once as the reference group.

These supplementary analyses produced patterns fully consistent with the results presented in the main text. Accordingly, no further model outputs are reported here; full results are provided in the Appendix (Tables 7–9).

Results on performance

The LMM was conducted with performance as the dependent variable. Random effects were specified for both classes and individual learners to account for the nested data structure. Fixed effects included engagement, measurement occasion (pre vs. post), intervention group (group S + SSI, S, SSI vs. control group), and their interactions. This model structure allows the examination of how engagement and intervention conditions influenced performance over time, whilst controlling for variability at the classroom and student levels.

Table 3 LMM with performance as the dependent variable, with class and students (ID) specified as random effects, and engagement, measurement occasion, intervention group, and their interaction as fixed effects (conditional $R^2 = 0.55$). Model specification: Performance $\sim 1 +$ Measurement + Engagement + Treatment + Measurement \times Treatment + (1 | class/ID)

	Variance	Std. dev.
Intercept (ID: class)	3.6729	1.9165
Intercept (class)	0.4018	0.6339
Residual	5.8997	2.4289

The variance components (Table 3) indicate that the majority of variance in performance occurred at the individual level rather than between classes. Specifically, student-level variance within classes ($\sigma^2 = 3.67$, $SD = 1.92$) exceeded class-level variance ($\sigma^2 = 0.40$, $SD = 0.63$), suggesting that differences between students were more pronounced than between classrooms. This finding indicates that individual differences between students account for most of the variance in performance, whereas class-level factors, such as teacher influence, play a comparatively smaller role. In other words, performance differences are primarily driven by students' individual characteristics and learning processes rather than by the specific class to which they belong. This suggests that student-level variables, such as performance or engagement, are likely to have a greater impact on learning outcomes than contextual classroom factors.

The LMM explained approximately 55% of the variance in performance ($R^2 = 0.55$), indicating a moderate to strong model fit.

The overall model effect was statistically significant, $b = 5.66$, $t(912) = 9.97$, $p < 0.001$, supporting the relevance of the included predictors. Both measurement occasion (post-test) and engagement were significant positive predictors of performance, indicating that performance improved following the intervention and that engagement was consistently associated with better outcomes. This means that students who reported higher levels of engagement also achieved better performance outcomes. Although the main effects for the treatment groups were non-significant, significant interaction effects emerged for groups SSI ($\beta = 0.42$, $p < 0.01$) and S ($\beta = 0.40$, $p < 0.01$). Students in these groups showed greater performance gains from pre- to post-test compared to the control group as shown in Table 4, suggesting that the intervention was particularly effective in these conditions. No significant interaction was found for group S + SSI in comparison to the control group. The two conditions that received only one treatment (SSI or S) were also significantly better compared to the S + SSI group. This indicates that students who experienced only one differentiation method, either scaffolding or SSI, showed significantly greater learning gains than those who received no differentiation or a combination of multiple differentiation methods.

Results on engagement

The LMM was conducted with engagement as the dependent variable. Random effects were specified for both classes and individual learners to account for the nested data structure. Fixed effects included measurement occasion (pre vs. post), performance, and intervention group (groups S + SSI, SSI, S vs.



Table 4 Standardised beta coefficients (with 95% confidence intervals) and *t*-values for the fixed effects in the LMM predicting performance. The model includes measurement occasion, treatment condition, engagement, and their interactions. Asterisks denote levels of statistical significance ($p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$). MT = measurement, T = Treatment, S = scaffolding, SSI = socio-scientific issues

	Std. beta [95% CI]	<i>t</i>
MT (measurement) [post]	0.64 [0.45, 0.84]***	6.43
T (treatment) [S + SSI]	0.16 [-0.13, 0.45]	1.09
T (treatment) [SSI]	-0.15 [-0.45, 0.15]	-0.99
T (treatment) [S]	-0.17 [-0.46, 0.12]	-1.14
Engagement	0.15 [0.09, 0.21]***	4.94
MT [post] × T [S + SSI]	0.06 [-0.19, 0.32]	0.50
MT [post] × T [SSI]	0.42 [0.15, 0.68]**	3.11
MT [post] × T [S]	0.40 [0.14, 0.65]**	3.02

Table 5 LMM with engagement as the dependent variable, with class and students (ID) specified as random effects, and measurement occasion, performance, and intervention group as fixed effects. Model specification: Engagement $\sim 1 + \text{Measurement} + \text{Performance} + \text{Treatment} + (1 | \text{class}/\text{ID})$

	Variance	Std. dev.
Intercept (ID: class)	0.23625	0.4861
Intercept (class)	0.02375	0.1541
Residual	0.14712	0.3836

control group). This model structure allows the examination of how performance and intervention conditions influenced engagement over time, whilst controlling for variability at the classroom and student levels. The comparison with the other intervention groups as a reference group can be found in the appendix, but further analyses confirm the results presented here. The variance components (Table 5) indicate again that the majority of variance in engagement occurred at the individual level ($\sigma^2 = 0.24$, $SD = 0.49$), with a notably smaller portion attributable to differences between classes ($\sigma^2 = 0.02$, $SD = 0.15$). The residual variance was 0.15 ($SD = 0.38$), suggesting that unmeasured individual factors still accounted for some variation.

The model (Table 6) explained approximately 65% of the variance in engagement ($R^2 = 0.65$), indicating a strong model fit. The overall model effect was statistically significant, $b = 2.15$, $t(915) = 22.76$, $p < 0.001$. Among the fixed effects, performance was a significant positive predictor of engagement ($\beta = 0.14$, $p < 0.001$), indicating that higher-performing students were also more engaged. However, neither measurement

Table 6 Standardised beta coefficients (with 95% confidence intervals) and *t*-values for the fixed effects in the LMM predicting engagement. Asterisks denote levels of statistical significance ($p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$). MT = measurement, T = treatment, S = scaffolding, SSI = socio-scientific issues

	Std. beta [95% CI]	<i>t</i>
MT (measurement) [post]	0.04 [-0.06, 0.13]	0.79
T (treatment) [S + SSI]	-0.31 [-0.64, 0.02]	-1.85
T (treatment) [SSI]	-0.20 [-0.53, 0.14]	-1.13
T (treatment) [S]	-0.24 [-0.57, 0.09]	-1.43
Performance	0.14 [0.07, 0.20]***	4.23

occasion ($\beta = 0.04$, $p = 0.43$) nor the treatment conditions showed significant effects on engagement. Although the coefficient for the S + SSI group was negative and approached significance ($\beta = -0.31$, $p = 0.065$), this effect did not reach conventional significance thresholds. These results indicate that students with higher performance levels also showed higher engagement. Students who were more successful in working through the material appeared to be more involved and motivated overall. At the same time, engagement remained relatively stable throughout the intervention and did not differ substantially between the treatment groups. The instructional methods applied, whether scaffolding, SSI, or their combination, did not produce measurable changes in students' engagement levels. The slightly lower engagement scores in the S + SSI group may reflect minor variations, but these differences were not statistically reliable.

Results on engagement subscales

In order to examine potential differential effects across the multidimensional construct of engagement, separate linear mixed models were conducted for emotional, cognitive, behavioural, agentic, and value-related engagement.

To examine the factorial structure of the engagement instrument, a confirmatory factor analysis (CFA) was conducted for the pre-test data (Fig. 12; Appendix). The hypothesised five-factor model comprising cognitive, emotional, behavioural, agentic, and value-related engagement showed an acceptable overall fit to the data, $\chi^2(199) = 1066.02$, $p < 0.001$, CFI = 0.846, TLI = 0.822, RMSEA = 0.081, 90% CI [0.076, 0.086], and SRMR = 0.078. Overall, the results provide partial support for the assumed multidimensional structure of the instrument. All factor loadings were statistically significant. The latent factor correlations were positive throughout, indicating that the five engagement dimensions were related but empirically distinguishable.

The model structure corresponded to that of the overall engagement analysis. Thus, the LMM was conducted with the respective engagement subscale as the dependent variable. Random effects were specified for classes and individual learners to account for the nested data structure. Fixed effects included measurement occasion (pre vs. post), performance, and intervention group (groups S + SSI, SSI, S vs. control group). This modelling approach allowed us to examine whether the effects observed for overall engagement were also reflected in the individual engagement dimensions.

Here, more effects are noticeable. These effects are quite low, and no significant differences between the intervention groups can be observed over time. The corresponding tables can be found in the appendix in "LMM Engagement Subscales" (Tables 10–19). However, the analysis reveals different trends over time across the engagement dimensions, regardless of the intervention group. Across the subscale models, no consistent main effects of the treatment conditions nor interaction effects between treatment and measurement occasion were observed. However, several engagement dimensions showed significant changes over time. Cognitive engagement increased slightly but significantly from pre- to post-measurement ($b = 0.10$, $t(919) = 2.16$, $p = 0.031$), and agentic engagement showed a



stronger positive change over time ($b = 0.28$, $t(916) = 7.65$, $p < 0.001$). In contrast, emotional engagement ($b = -0.17$, $t(918) = -3.99$, $p < 0.001$) and behavioural engagement ($b = -0.11$, $t(915) = -2.84$, $p = 0.005$) decreased significantly across the measurement occasions. No significant change over time was observed for value-related engagement ($b = 0.07$, $t(915) = 1.38$, $p = 0.167$).

Taken together, these results indicate that the different dimensions of engagement developed in distinct directions over the course of the intervention. While students reported slight increases in cognitive and agentic engagement, suggesting greater mental effort and more active involvement in shaping their learning process, emotional and behavioural engagement declined over time. This pattern may reflect that students became more cognitively involved in the learning tasks while simultaneously experiencing decreasing affective enthusiasm or persistence during the longer intervention period. As these developments occurred across all intervention groups and in opposite directions, they likely offset each other when engagement was analysed as an overall construct.

Discussion

The results of this study provide valuable insights into how SSI and scaffolding influence student performance and engagement in chemistry education. The research question asked how these two instructional strategies, individually and in combination, influence student affective and cognitive outcomes, particularly in the context of ESD. The hypotheses derived from the literature were: (1) engagement is positively influenced by SSI, (2) performance is positively influenced by scaffolding, and (3) both engagement and performance are positively influenced by the combined use of SSI and scaffolding.

With reference to the hypotheses formulated, only the second hypothesis can be confirmed with the data. Students who received scaffolding (Group S) demonstrated significant gains in performance compared to the control group (Group CG). The first and third hypotheses, in contrast, were not supported: neither SSI (Group SSI) alone nor the combination of SSI and scaffolding (Group S + SSI) produced statistically significant increases in engagement, and the combined intervention did not yield additive effects on performance. However, a positive influence of SSI on subject knowledge was observed.

With regard to Hypothesis 1, it is necessary to examine the construct of engagement first. Engagement emerged as a differentiated construct in the present study, revealing a more nuanced pattern than suggested by the overall engagement scale alone. While no significant changes in global engagement were observed across measurement occasions or intervention groups, the analyses of the engagement subscales indicate divergent developmental trajectories over time. These findings underline the importance of conceptualising engagement as a multidimensional and dynamic construct rather than a unidimensional outcome (Fredricks *et al.*, 2004; Reeve, 2013).

Consistent with prior research, performance was a significant positive predictor of overall engagement. Higher-performing

students reported higher engagement levels, supporting the well-established reciprocal relationship between achievement and engagement (Fredricks *et al.*, 2004; Azevedo, 2015). Learners who experience success and mastery are more likely to invest effort, persist in challenging tasks and remain cognitively involved (Bandura, 1997; Potvin and Hasin, 2014). This association was particularly evident in the subscale of cognitive engagement, where gains over time were significantly correlated with performance gains. Cognitive engagement, as operationalised in this study, explicitly targets mental effort, strategy use, and deep processing, which are closely aligned with performance-related learning outcomes. These results are in line with research on self-regulated learning, showing that strategic cognitive engagement is both a predictor and a consequence of academic success (Pintrich, 2000; Zimmerman, 2008).

Beyond cognitive engagement, agentic engagement also showed a significant positive increase over time, independent of intervention condition. Agentic engagement refers to students' proactive contribution to the learning process, such as asking questions, expressing preferences, and influencing instructional activities (Reeve and Tseng, 2011). The parallel increase in cognitive and agentic engagement suggests that the learning environment fostered aspects of self-regulation and learner autonomy. Even though the instructional design did not explicitly target agentic behaviours, the largely self-directed, tablet-based learning format encouraged students to take greater responsibility for their learning processes. Previous studies have shown that environments that allow choice, autonomy, and independent pacing can promote agentic engagement, particularly when learners perceive themselves as competent (Reeve, 2013; Jang *et al.*, 2016).

In contrast, emotional and behavioural engagement showed significant negative trends over time, although numerous collaborative tasks were included. With reference to findings from research on situational interest, emotional engagement, which encompasses interest, enjoyment, and feelings of belonging, tends to decrease when instructional contexts remain structurally stable over extended periods, like working self-regulated on the LL (Renninger and Hidi, 2015; Rotgans and Schmidt, 2018).

At the same time, there is a close relation between behavioural and emotional engagement, as both contribute to learners' endurance and willingness to remain actively involved in learning activities (Li *et al.*, 2023). From this perspective, the negative trends in emotional and behavioural engagement appear less surprising and do not necessarily indicate a failure of the instructional approach, but rather highlight the temporal dynamics of engagement during longer interventions (Fredricks *et al.*, 2004).

However, no significant effects were observed for value-related engagement. This suggests that students' perceptions of the personal relevance or importance of chemistry remained relatively stable over time. Value-related engagement is often considered a more enduring, belief-like component of motivation that is less susceptible to short-term instructional influences (Eccles and Wigfield, 2002; Fechner, 2009; Krapp and Prenzel, 2011). Although context-based and SSI-oriented instructions have been shown to enhance perceived relevance



in some studies (Badeo and Duque, 2022; Klaver *et al.*, 2023), such effects are not consistently replicated and may depend strongly on learners' prior interests and experiences. In the present study, the absence of change in value-related engagement should therefore not be interpreted as evidence that the SSI were generally not meaningful or relevant to students. Rather, research suggests that value-related and interest-related aspects of engagement develop more slowly and require sustained support over time (Hattie, 2009; van de Pol *et al.*, 2015). In particular, the development of stable value beliefs is associated with repeated opportunities for reflection and personal relevance (Wigfield and Eccles, 2000).

Importantly, no differential effects between the intervention groups emerged across any engagement dimension. Neither SSI, scaffolding, nor their combination led to distinct engagement trajectories. Previous research results confirm that scaffolding primarily addresses cognitive rather than affective processes, which is in line with results in our study (van de Pol *et al.*, 2015; Janson *et al.*, 2020; Faber *et al.*, 2024). While scaffolding has been shown to support motivation and engagement, most implementations focus on cognitive dimensions such as problem-solving or conceptual understanding (Belland *et al.*, 2017). To systematically influence emotional or value-related engagement, scaffolding would need to explicitly incorporate motivational prompts, reflective activities, or opportunities for personal meaning-making (Reeve, 2013).

In parallel with Hypothesis 2, the positive impact of scaffolding on performance is consistent with the literature (*e.g.*, Belland *et al.*, 2017). By providing adaptive support tailored to students' levels of understanding, scaffolding allows learners to confront challenging content at their own pace and build confidence in problem solving. In line with the results of Hammond and Gibbons (2005), this study also suggests that differentiated scaffolds (strategic, conceptual, and further thinking) support students of varying proficiency in accessing the material and making progress. The present study did not include usage-tracking data on individual scaffold interaction. Consequently, it cannot be determined which specific scaffolds were accessed by which students, nor to what extent individual differences in usage contributed to performance outcomes.

Notably, the variance components from the LMM revealed that most of the variability in performance was attributable to differences between individual students rather than between classes. This suggests that it is individual learner characteristics, such as prior knowledge, engagement, or learning strategies, that predominantly shape performance outcomes. This is also supported by the research reviewed so far. Belland *et al.* (2017) demonstrated that adaptive scaffolds provide particularly effective support for learners' cognitive processes. Similarly, Güth and van Vorst (2023) showed that affective factors can be fostered through differentiated context designs, as learners choose different contexts for their learning in chemistry.

The results show that not only the use of scaffolding in group S, but also the integration of SSI in group SSI, positively influenced student performance, as both groups demonstrated significantly greater gains than the control group. Consistent

with the findings of the present study, prior research indicates that SSI-based instruction can positively influence students' conceptual understanding. For instance, Sadler *et al.* (2016) showed that embedding scientific content in socio-scientific contexts promotes deeper engagement with disciplinary ideas. This provides a possible explanation for the positive effects observed in the present study.

The combination of SSI and scaffolding did not yield the additive or synergistic effect that was initially anticipated with Hypothesis 3. Students in the combined group (S + SSI) did not outperform those in the single-treatment groups, despite experiencing both real-world context and scaffolds. Moreover, students in the combined group (Group S + SSI) did not differ significantly from students in the control group in post-test performance. The analysis revealed no significant interaction effect between SSI and scaffolding on performance, meaning that using both strategies together did not significantly support performance beyond the gains achieved by each strategy alone. However, the present data do not allow for a conclusive explanation of this finding. While the lack of additional effects in the combined condition may be related to factors such as cognitive load (Schnotz and Rasch, 2005) or insufficient alignment of instructional elements, these interpretations remain speculative and cannot be substantiated based on the current data. But, these findings indicate that combining multiple differentiation approaches does not necessarily lead to additional learning gains and may even reduce effectiveness if the instructional design is not carefully coordinated.

In summary, the present findings suggest that differentiated digital learning materials in the LL can foster self-regulated dimensions of engagement, particularly cognitive and agentic engagement, regardless of the specific instructional condition. At the same time, affective and persistence-related aspects of engagement appear more difficult to sustain over longer intervention periods and may require targeted instructional support. These results underscore the necessity of addressing engagement as a multidimensional construct in both research and instructional design, and caution against drawing conclusions based solely on aggregated engagement measures.

Taken together, the findings indicate that both SSI and scaffolding approaches can enhance student performance in chemistry while providing meaningful opportunities to integrate sustainability into science education. At the same time, the results suggest that combining these strategies does not automatically lead to additional learning gains.

Limitations

Some limitations of the present study should be considered when interpreting the results. First, the intervention was implemented as a quasi-experimental study in real classrooms rather than a fully randomized controlled experiment. The six-week duration and the natural classroom setting meant that not all external variables, such as differences in teaching style or classroom dynamics, could be strictly controlled. While this limits the internal validity of



causal claims, it also offers the advantage of demonstrating how the approach works in a real-world educational context, thus enhancing the ecological validity of the findings. Even though our LMM showed that the class factor had little influence.

Second, the sample was limited to students from academically oriented secondary schools (Gymnasium) in Germany. As a result, the findings may not fully represent the diversity of learners in other educational tracks or cultural contexts. The focused sample did allow for consistency in instructional conditions and comparability within this school type, but it narrows the scope of generalization. Future studies should include a broader range of school contexts and student backgrounds to examine whether learners with different academic profiles respond differently to SSI-based and scaffolded instruction.

It should be considered that students in all conditions worked in a digital, self-directed learning environment, which may have introduced a novelty effect initially. However, given the six-week duration of the intervention and students' prior experience with tablets in regular classroom practice, such effects likely diminished over time. As all groups used the same learning environment, potential novelty effects would have influenced all conditions equally. Nevertheless, the design of the learning environment may have affected learning behaviour independently of the instructional conditions.

Additionally, no data were collected on students' cognitive load during the learning activities. It is conceivable that the combination of SSI contexts and scaffolding, while intended to enhance learning, might have introduced a higher cognitive demand on students. Without measures of cognitive load, it is not possible to determine whether the lack of an additive effect in the combined condition was partly due to students experiencing overload or difficulty in processing the materials. Incorporating cognitive load assessments in future studies would provide valuable insights into how students are managing the complex tasks and whether adjustments to the scaffolding design could alleviate any undue cognitive burden.

Also, no tracking or learning analytics data were collected to document students' actual use of the optional scaffolds. Consequently, it remains unclear which specific scaffold elements were accessed by individual learners and to what extent differences in usage patterns may have contributed to the observed performance outcomes. Furthermore, the study did not include a detailed qualitative analysis of how students interacted with the learning materials. A finer-grained look at student behaviours, for instance, through classroom observations, student interviews, or analysis of student work and help-seeking patterns, would be a great complement to the quantitative findings. Such qualitative insights could illuminate the mechanisms behind the observed outcomes and inform improvements in the design and implementation of SSI-based, scaffolded learning units.

It should also be noted that not all model fit indices in the (longitudinal) CFAs indicated an optimal fit. Although the overall model fit was within an acceptable range, the results should be interpreted with some caution. Nevertheless, the theoretically grounded factor structure, significant factor

loadings, and evidence for measurement invariance across time support the use of the model for longitudinal analyses.

In summary, while this quasi-experimental study offers valuable evidence of the potential of SSI and scaffolding in a real classroom setting to implement ESD, these limitations highlight the need for cautious interpretation and guide the directions for future research. Addressing these limitations in subsequent studies, through more varied samples, longer intervention periods, additional measurement of cognitive and behavioural processes, and mixed-method approaches will help in building a more robust understanding of how to effectively differentiate science instruction for both cognitive and affective gains. Nevertheless, despite these limitations, the present study provides meaningful insights into how learners with different performance and engagement respond to SSI-based and scaffolded instruction. It demonstrates that differentiated instruction designs can yield substantial learning gains for ESD even in heterogeneous classrooms and offers an empirically grounded way for developing adaptive science instruction that more effectively supports diverse learners. In addition, it is overall a successful material for ESD, which is adaptively designed for use in everyday school life and tailored to the needs of learners.

Implications for classroom instruction

The contribution of this work lies in its combined focus on cognitive and affective outcomes within an ESD-oriented chemistry learning environment. In contrast to earlier ESD-based interventions that primarily emphasized competences and decision-making skills development (Cebrián *et al.*, 2020; Olsson *et al.*, 2022; Riess *et al.*, 2022), this study explicitly considered an affective dimension of learning by examining student engagement alongside performance. Examining scaffolding and SSI strategies within a real-world classroom context further adds to the significance of the findings, as it shows that such innovations can be practically implemented at schools. The learning unit developed for this study is empirically validated and can be adopted by educators to enrich the relevance of chemistry instruction with sustainability-related content using proven differentiation methods.

The findings offer cautious implications for classroom instruction, particularly when integrating ESD into practice. Both SSI and scaffolding independently supported students' performance in the present study. Embedding chemistry content in socio-scientific contexts can help connect scientific concepts to real-world challenges. At the same time, scaffolding appears to support learners in dealing with complex tasks by providing adaptive guidance that accommodates different levels of prior knowledge and problem-solving ability.

Furthermore, the results point to substantial variability at the individual learner level, suggesting that students respond differently to the instructional approach. This variability points to the need for differentiated instructional strategies in heterogeneous classrooms, where flexible support structures enable students to engage with tasks at varying levels of complexity.



Such approaches may help address differences in learners' needs and learning processes.

Although SSI and scaffolding each supported performance, their combination did not result in additional learning gains. This suggests that different instructional elements need to be carefully aligned, so that they complement rather than over-complicate the learning process.

The findings related to engagement further underline the importance of considering engagement as a multidimensional construct in instructional design. As the present results show that different engagement dimensions can develop in distinct directions over time, a revised version of the LL on organic chemistry should address not only cognitive but also emotional and behavioural aspects of engagement. Incorporating opportunities for peer discussion, collaborative problem-solving, or reflective activities that encourage students to connect SSI with their own perspectives and experiences could be strengthened to address this aspect.

From a practical perspective, to deliberately influence affective outcomes, scaffolding strategies would need to explicitly target emotional and motivational components, for example by integrating reflection prompts or opportunities for self-expression. This distinction is crucial for future instructional design and research: it may not be sufficient to assume that scaffolds automatically support engagement without addressing affective dimensions directly. For instance, incorporating more reflective elements, peer discussion, or real-world relevance could strengthen the emotional dimension of engagement. For future research, it will be important to explore how longer interventions, repeated exposure, or explicitly scaffolding strategies might also influence engagement. Moreover, identifying how individual learner characteristics, such as prior interest or perceived competence, mediate the relationship between engagement and performance could offer valuable insights for the design of adaptive learning environments.

Overall, the learning unit for organic chemistry, incorporating SSI contexts and scaffolding strategies, represents a promising approach for integrating sustainability-related topics into chemistry education. However, their effectiveness appears to depend, at least in part, on careful instructional design, particularly with regard to the balance between cognitive support, social interaction, and opportunities for reflection.

Ethical considerations

The study was approved by the Ethics Committee of the Interdisciplinary Center for Educational Research (IZfB) at the University of Duisburg-Essen and complied with all ethical guidelines for research on humans. Prior to the data collection, all participants were informed about the data collection and, if desired, confirmation was obtained from their legal guardians, ensuring their voluntary participation and the confidentiality of their responses. All data was collected anonymously to protect participants' privacy, and no identifying information was included in the analysis or reporting of results.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data generated and analysed in this study involve human participants and are subject to ethical and data protection restrictions. In line with the conditions of informed consent and the institutional data protection policy, the data cannot be shared or made publicly available. Therefore, the data supporting the findings of this study are not available.

Appendix

Descriptive statistics for engagement subscales

Cognitive engagement

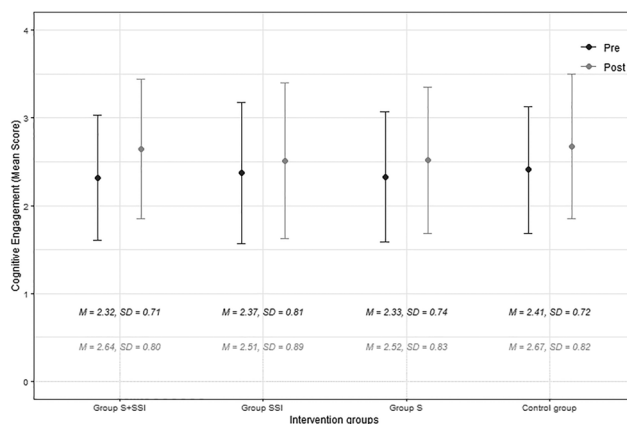


Fig. 7 Descriptive statistics for cognitive engagement (mean score) with means (M) and standard deviation (SD) for all intervention groups pre and post; S = scaffolding, SSI = socio-scientific issues.

Emotional engagement

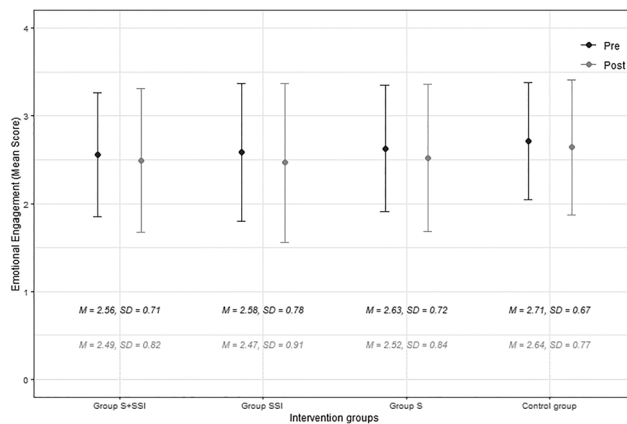


Fig. 8 Descriptive statistics for emotional engagement (mean score) with means (M) and standard deviation (SD) for all intervention groups pre and post; S = scaffolding, SSI = Socio-scientific issues.



Behavioural engagement

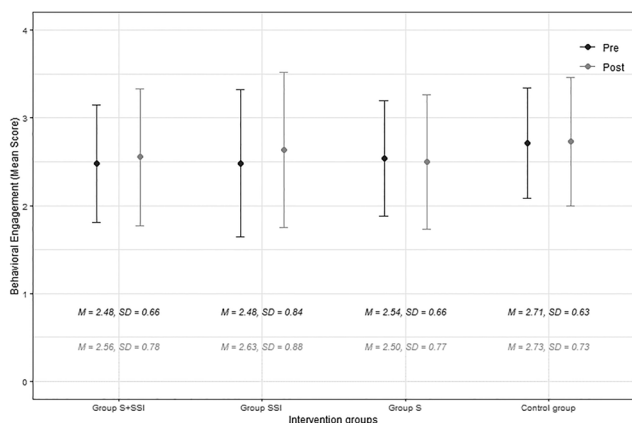


Fig. 9 Descriptive statistics for emotional engagement (mean score) with means (M) and standard deviation (SD) for all intervention groups pre and post; S = scaffolding, SSI = socio-scientific issues.

Agentive engagement

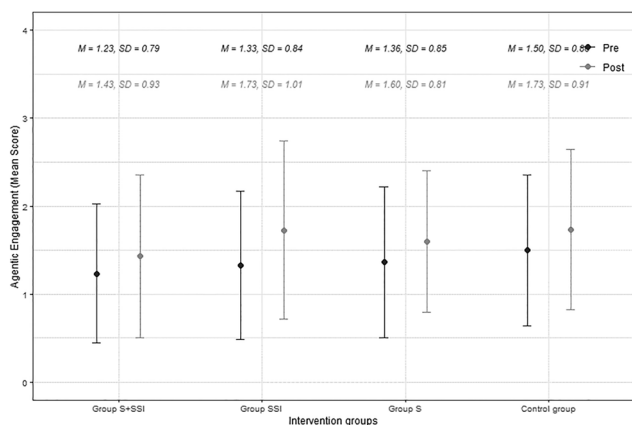


Fig. 10 Descriptive statistics for emotional engagement (mean score) with means (M) and standard deviation (SD) for all intervention groups pre and post; S = scaffolding, SSI = socio-scientific issues.

Value-related engagement

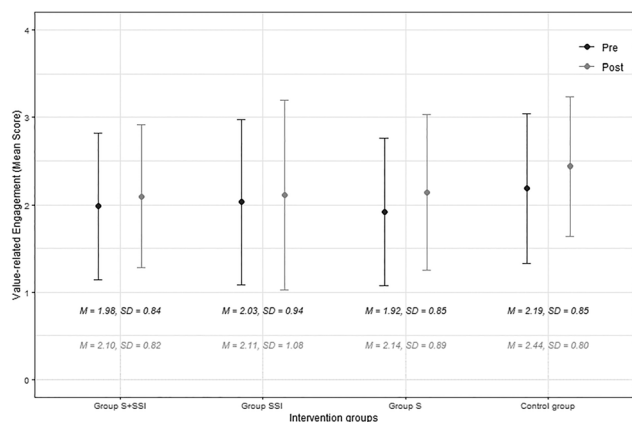


Fig. 11 Descriptive statistics for emotional engagement (mean score) with means (M) and standard deviation (SD) for all intervention groups pre and post; S = scaffolding, SSI = socio-scientific issues.

LMM performance; other reference groups (RG)

Reference group: S + SSI

Table 7 Standardised beta coefficients (with 95% confidence intervals) and *t*-values for the fixed effects in the LMM predicting performance. The model includes measurement occasion, treatment condition, engagement, and their interactions. Asterisks denote levels of statistical significance ($p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$). RG = S + SSI

	Std. beta [95% CI]	<i>t</i>
MT (measurement) [post]	0.71, [0.55, 0.87]***	8.64
T (treatment) [CG]	-0.16, [-0.45, 0.13]	-1.09
T (treatment) [SSI]	-0.31, [-0.58, -0.04]*	-2.44
T (treatment) [S]	-0.33, [-0.59, -0.06]*	-2.23
Engagement	0.15, [0.09, 0.21]***	4.94
MT [post] × T [CG]	-0.06, [-0.32, 0.19]	-0.50
MT [post] × T [SSI]	0.35, [0.11, 0.59]**	2.91
MT [post] × T [S]	0.33, [0.10, 0.56]**	2.80

Reference group: SSI

Table 8 Standardised beta coefficients (with 95% confidence intervals) and *t*-values for the fixed effects in the LMM predicting performance. The model includes measurement occasion, treatment condition, engagement, and their interactions. Asterisks denote levels of statistical significance ($p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$). RG = SSI

	Std. beta [95% CI]	<i>t</i>
MT (measurement) [post]	1.06, [0.89, 1.23]***	11.93
T (treatment) [S + SSI]	0.31, [0.04, 0.58]*	2.23
T (treatment) [CG]	0.15, [-0.15, 0.45]	0.99
T (treatment) [S]	-0.02, [-0.29, 0.26]	-0.13
Engagement	0.15, [0.09, 0.21]***	4.94
MT [post] × T [S + SSI]	-0.35, [-0.59, -0.11]**	-2.91
MT [post] × T [CG]	-0.42, [-0.68, -0.15]**	-3.11
MT [post] × T [S]	-0.02, [-0.26, 0.22]	-0.16

Reference group: S

Table 9 Standardised beta coefficients (with 95% confidence intervals) and *t*-values for the fixed effects in the LMM predicting performance. The model includes measurement occasion, treatment condition, engagement, and their interactions. Asterisks denote levels of statistical significance ($p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$). RG = S

	Std. beta [95% CI]	<i>t</i>
MT (measurement) [post]	1.04, [0.87, 1.21]***	12.21
T (treatment) [SSI]	0.02, [-0.26, 0.29]	0.13
T (treatment) [S + SSI]	0.33, [0.06, 0.59]*	2.44
T (treatment) [CG]	0.17, [-0.12, 0.46]	1.14
Engagement	0.15, [0.09, 0.21]***	4.94
MT [post] × T [SSI]	0.02, [-0.22, 0.26]	0.16
MT [post] × T [S + SSI]	-0.33, [-0.56, -0.10]**	-2.80
MT [post] × T [CG]	-0.40, [-0.65, -0.14]**	-3.02



LMM engagement subscales

Cognitive engagement

Table 10 LMM with cognitive engagement as the dependent variable, with students (ID) specified as a random effect, and performance and measurement occasion as fixed effects (conditional $R^2 = 0.46$). Model specification: Cognitive Engagement $\sim 1 + \text{Measurement} + \text{Performance} + (1|ID)$

	Variance	Std. dev.
Intercept (ID)	0.2602	0.5101
Residual	0.3491	0.5908

Table 11 Standardised beta coefficients (with 95% confidence intervals) and t -values for the fixed effects in the LMM predicting cognitive engagement. The model includes measurement occasion and performance. Asterisks denote levels of statistical significance ($p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$)

	Std. beta [95% CI]	t
MT (measurement) [post]	0.12, [0.01, 0.24]*	2.16
Performance	0.20, [0.13, 0.27]***	5.67

Emotional engagement

Table 12 LMM with emotional engagement as the dependent variable, with students (ID) and class specified as random effects, and performance and measurement occasion as fixed effects (conditional $R^2 = 0.53$). Model specification: Emotional Engagement $\sim 1 + \text{Measurement} + \text{FW} + (1|\text{class}/ID)$

	Variance	Std. dev.
Intercept (ID: class)	0.28929	0.5379
Intercept (class)	0.03334	0.1826
Residual	0.28982	0.5384

Table 13 Standardised beta coefficients (with 95% confidence intervals) and t -values for the fixed effects in the LMM predicting emotional engagement. The model includes measurement occasion and performance. Asterisks denote levels of statistical significance ($p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$)

	Std. beta [95% CI]	t
MT (measurement) [post]	-0.22, [-0.33, -0.11]***	-3.99
Performance	0.10, [0.03, 0.17]**	2.96

Behavioural engagement

Table 14 LMM with behavioural engagement as the dependent variable, with students (ID) and class specified as random effects, and performance and measurement occasion as fixed effects. (conditional $R^2 = 0.61$) Model specification: Behavioural Engagement $\sim 1 + \text{Measurement} + \text{Performance} + \text{Treatment} + (1|\text{class}/ID)$

	Variance	Std. dev.
Intercept (ID: class)	0.29049	0.5390
Intercept (class)	0.02332	0.1527
Residual	0.21759	0.4665

Table 15 Standardised beta coefficients (with 95% confidence intervals) and t -values for the fixed effects in the LMM predicting behavioural engagement. The model includes measurement occasion, performance and treatment. Asterisks denote levels of statistical significance ($p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$)

	Std. beta [95% CI]	t
MT (measurement) [post]	-0.14, [-0.24, -0.04]**	-2.84
Performance	0.18, [0.11, 0.24]***	5.42
T (treatment) [S + SSI]	-0.33, [-0.63, -0.02]*	-2.12
T (treatment) [SSI]	-0.21, [-0.53, 0.10]	-1.33
T (treatment) [S]	-0.31, [-0.61, -7.15 $\times 10^{-4}$]*	-1.97

Agentic engagement

Table 16 LMM with agentic engagement as the dependent variable, with students (ID) and class specified as random effects, and performance and measurement occasion as fixed effects (conditional $R^2 = 0.61$). Model specification: Agentic Engagement $\sim 1 + \text{Measurement} + \text{Treatment} + (1|\text{class}/ID)$

	Variance	Std. dev.
Intercept (ID: class)	0.41567	0.6447
Intercept (class)	0.03614	0.1901
Residual	0.31419	0.5605

Table 17 Standardised beta coefficients (with 95% confidence intervals) and t -values for the fixed effects in the LMM predicting agentic engagement. The model includes measurement occasion and treatment. Asterisks denote levels of statistical significance ($p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$)

	Std. beta [95% CI]	t
MT (measurement) [post]	0.32, [0.24, 0.40]***	7.65
T (treatment) [S + SSI]	-0.32, [-0.64, -9.13 $\times 10^{-3}$]*	-2.02
T (treatment) [SSI]	-0.12, [-0.44, 0.19]	-0.77
T (treatment) [S]	-0.08, [-0.40, 0.25]	-0.47

Value-related engagement

Table 18 LMM with value-related engagement as the dependent variable, with students (ID) and class specified as random effects, and performance and measurement occasion as fixed effects (conditional $R^2 = 0.50$). Model specification: value-related Engagement $\sim 1 + \text{Measurement} + \text{Performance} + \text{Treatment} + (1|\text{class}/ID)$

	Variance	Std. dev.
Intercept (ID: class)	0.3555	0.5962
Intercept (class)	0.0286	0.1691
Residual	0.4112	0.6413

Table 19 Standardised beta coefficients (with 95% confidence intervals) and t -values for the fixed effects in the LMM predicting value-related engagement. The model includes measurement occasion and treatment. Asterisks denote levels of statistical significance ($p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$)

	Std. beta [95% CI]	T
MT (measurement) [post]	0.08, [-0.03, 0.19]	7.65
Performance	0.14, [0.08, 0.21]***	4.11
T (treatment) [S + SSI]	-0.30, [-0.59, -0.01]*	-2.05
T (treatment) [SSI]	-0.30, [-0.60, -4.02 $\times 10^{-3}$]*	-1.99
T (treatment) [S]	-0.32, [-0.61, -0.03]*	-2.19



Confirmatory factor analysis (CFA)

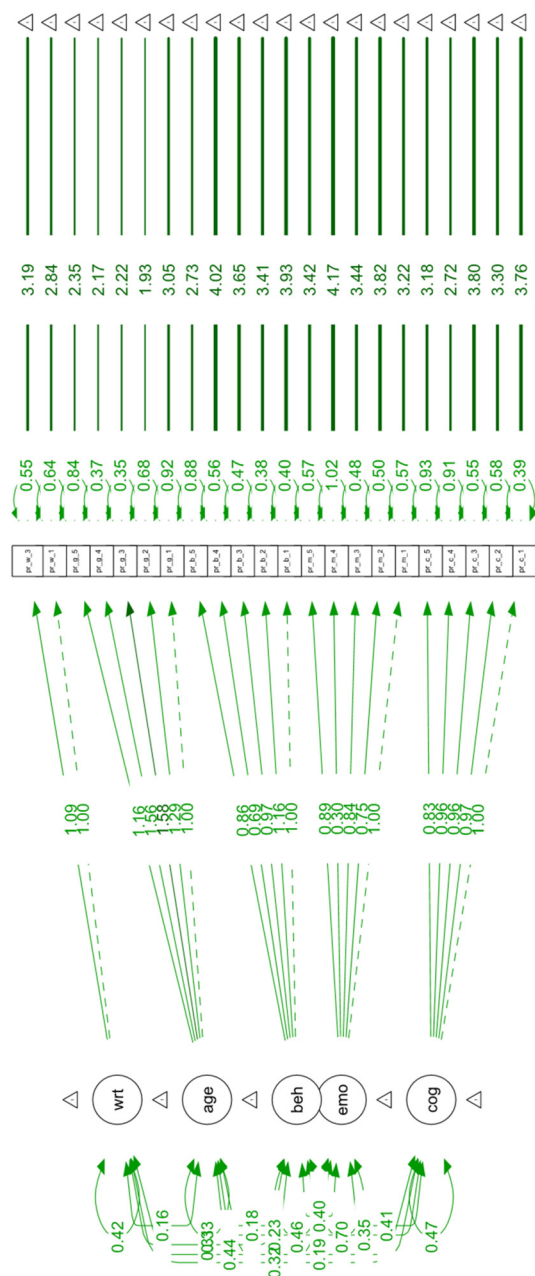


Fig. 12 Confirmatory factor analysis for the pre-test data. The five-factor model comprising cognitive, emotional, behavioural, agentive, and value-related engagement.

References

- Acosta-Gonzaga, E. and Ramirez-Arellano, A., (2022), Scaffolding Matters? Investigating Its Role in Motivation, Engagement and Learning Achievements in Higher Education, *Sustainability*, **14**(20), 1–17.
- Appleton, J. J., Christenson, S. L. and Furlong, M. J., (2008), Student engagement with school: Critical conceptual and

methodological issues of the construct, *Psychol. Sch.*, **45**(5), 369–386, DOI: [10.1002/pits.20303](https://doi.org/10.1002/pits.20303).

- Avargil, S., Kohen, Z. and Dori, Y. J., (2020), Trends and perceptions of choosing chemistry as a major and a career, *Chem. Educ. Res. Pract.*, **21**(2), 668–684, DOI: [10.1039/C9RP00158A](https://doi.org/10.1039/C9RP00158A).
- Azevedo, R., (2015), Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues, *Educ. Psychol.*, **50**(1), 84–94, DOI: [10.1080/00461520.2015.1004069](https://doi.org/10.1080/00461520.2015.1004069).
- Azevedo, R. and Hadwin, A. F., (2005), Scaffolding Self-regulated Learning and Metacognition – Implications for the Design of Computer-based Scaffolds, *Instruct. Sci.*, **33**(5–6), 367–379, DOI: [10.1007/s11251-005-1272-9](https://doi.org/10.1007/s11251-005-1272-9).
- Azevedo, R., Cromley, J. G. and Seibert, D., (2004), Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemp. Educ. Psychol.*, **29**(3), 344–370, DOI: [10.1016/j.cedpsych.2003.09.002](https://doi.org/10.1016/j.cedpsych.2003.09.002).
- Badeo, J. M. and Duque, D. A., (2022), The effect of Socio-Scientific Issues (SSI) in teaching science: A meta-analysis study, *J. Technol. Sci. Educ.*, **12**(2), 291, DOI: [10.3926/jotse.1340](https://doi.org/10.3926/jotse.1340).
- Bandura, A., (1997), *Self-efficacy: The exercise of control*. *Self-efficacy: The exercise of control*, W H Freeman/Times Books/Henry Holt & Co.
- Barnes, S. A., Mallinckrodt, C. H., Lindborg, S. R. and Carter, M. K., (2008), The impact of missing data and how it is handled on the rate of false-positive results in drug development, *Pharm. Stat.*, **7**(3), 215–225, DOI: [10.1002/pst.310](https://doi.org/10.1002/pst.310).
- Belland, B. R., Kim, C. and Hannafin, M. J., (2013), A Framework for Designing Scaffolds That Improve Motivation and Cognition, *Educ. Psychol.*, **48**(4), 243–270, DOI: [10.1080/00461520.2013.838920](https://doi.org/10.1080/00461520.2013.838920).
- Belland, B. R., Walker, A. E., Kim, N. J. and Lefler, M., (2017), Synthesizing Results From Empirical Research on Computer-Based Scaffolding in STEM Education: A Meta-Analysis, *Rev. Educ. Res.*, **87**(2), 309–344, DOI: [10.3102/0034654316670999](https://doi.org/10.3102/0034654316670999).
- Belland, B. R., Lee, E., Zhang, A. Y. and Kim, C., (2022), Characterizing the most effective scaffolding approaches in engineering and technology education: A clustering approach, *Comput. Appl. Eng. Educ.*, **30**(6), 1795–1812, DOI: [10.1002/cae.22556](https://doi.org/10.1002/cae.22556).
- Benjamini, Y. and Hochberg, Y., (1995), Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *J. R. Stat. Soc. Ser. B: Stat. Methodol.*, **57**(1), 289–300, DOI: [10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x).
- Bennett, D., Power, A., Thomson, C., Mason, B. and Bartleet, B.-L., (2016), Reflection for learning, learning for reflection: Developing Indigenous competencies in higher education, *J. Univ. Teach. Learn. Pract.*, **13**(2), 99–118, DOI: [10.53761/1.13.2.7](https://doi.org/10.53761/1.13.2.7).
- Bennett, J. and Lubben, F., (2006), Context-based Chemistry: The Salters approach, *Int. J. Sci. Educ.*, **28**(9), 999–1015, DOI: [10.1080/09500690600702496](https://doi.org/10.1080/09500690600702496).
- Bond, T., (2015), *Applying the Rasch Model*, Routledge.



- Bondie, R. S., Dahnke, C. and Zusho, A., (2019), How Does Changing “One-Size-Fits-All” to Differentiated Instruction Affect Teaching? *Rev. Res. Educ.*, **43**(1), 336–362, DOI: [10.3102/0091732X18821130](https://doi.org/10.3102/0091732X18821130).
- Broman, K., Bernholt, S. and Parchmann, I., (2018), Using model-based scaffolds to support students solving context-based chemistry problems, *Int. J. Sci. Educ.*, **40**(10), 1176–1197, DOI: [10.1080/09500693.2018.1470350](https://doi.org/10.1080/09500693.2018.1470350).
- Brooke J., (1995), SUS: A quick and dirty usability scale, *Usability Eval. Ind.*, 189.
- Brown, T. A., (2015), *Confirmatory factor analysis for applied research (Second edition)*, *Methodology in the social sciences*, The Guilford Press.
- Bulte, A. M. W., Westbroek, H. B., Jong, O. de and Pilot, A., (2006), A Research Approach to Designing Chemistry Education using Authentic Practices as Contexts, *Int. J. Sci. Educ.*, **28**(9), 1063–1086, DOI: [10.1080/09500690600702520](https://doi.org/10.1080/09500690600702520).
- Burmeister, M., Rauch, F. and Eilks, I., (2012), Education for Sustainable Development (ESD) and chemistry education, *Chem. Educ. Res. Pract.*, **13**(2), 59–68, DOI: [10.1039/C1RP90060A](https://doi.org/10.1039/C1RP90060A).
- Çalık, M. and Wiyarsi, A., (2021), A systematic review of the research papers on chemistry-focused socio-scientific issues, *J. Baltic Sci. Educ.*, **20**(3), 360–372, DOI: [10.33225/jbse/21.20.360](https://doi.org/10.33225/jbse/21.20.360).
- Cebrián, G., Junyent, M. and Mulà, I., (2020), Competencies in Education for Sustainable Development: Emerging Teaching and Research Developments, *Sustainability*, **12**(2), 579, DOI: [10.3390/su12020579](https://doi.org/10.3390/su12020579).
- Celik, K. N., (2022), *Entwicklung Von Chemischem Fachwissen in der Sekundarstufe I: Validierung Einer Learning Progression Für Die Basiskonzepte Struktur der Materie Chemische Reaktion und Energie Im Kompetenzbereich Umgang Mit Fachwissen. Studien zum Physik- und Chemielernen: [Development of Chemical Expertise in Secondary Level I: Validation of a Learning Progression for the Basic Concepts of Structure of Matter, Chemical Reactions, and Energy in the Competence Area of Handling Expertise. Studies on Physics and Chemistry Learning:] v.325*, Logos Verlag Berlin.
- Chen, F. F., (2007), Sensitivity of goodness of fit indexes to lack of measurement invariance, *Struct. Eq. Model.*, **14**(3), 464–504, DOI: [10.1080/10705510701301834](https://doi.org/10.1080/10705510701301834).
- Chernikova, O., Sommerhoff, D., Stadler, M., Holzberger, D., Nickl, M., Seidel, T., Kasneci, E., Küchemann, S., Kuhn, J., Fischer, F. and Heitzmann, N., (2025), Personalization through adaptivity or adaptability? A meta-analysis on simulation-based learning in higher education, *Educ. Res. Rev.*, **46**, 100662, DOI: [10.1016/j.edurev.2024.100662](https://doi.org/10.1016/j.edurev.2024.100662).
- Cohen, J., (2013), *Statistical Power Analysis for the Behavioral Sciences*, Routledge.
- Deunk, M. I., Smale-Jacobse, A. E., Boer, H., de Doolaard, S. and Bosker, R. J., (2018), Effective differentiation Practices: A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education, *Educ. Res. Rev.*, **24**, 31–54, DOI: [10.1016/j.edurev.2018.02.002](https://doi.org/10.1016/j.edurev.2018.02.002).
- Eccles, J. S. and Wigfield, A., (2002), Motivational beliefs, values, and goals, *Annu. Rev. Psychol.*, **53**, 109–132, DOI: [10.1146/annurev.psych.53.100901.135153](https://doi.org/10.1146/annurev.psych.53.100901.135153).
- Engeln K., (2004), Schülerlabors: Authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken, Dissertation, Science Labs for Students: Authentic, Engaging Learning Environments as a Way to Spark Interest in Science and Technology, Berlin.
- Faber, T. J. E., Dankbaar, M. E. W., van den Broek, W. W., Bruinink, L. J., Hogeveen, M. and van Merriënboer, J. J. G., (2024), Effects of adaptive scaffolding on performance, cognitive load and engagement in game-based learning: A randomized controlled trial, *BMC Med. Educ.*, **24**(1), 943, DOI: [10.1186/s12909-024-05698-3](https://doi.org/10.1186/s12909-024-05698-3).
- Fechner, S., (2009), *Effects of context oriented learning on student interest and achievement in chemistry education. Studien zum Physik- und Chemielernen: Vol. 95*, Logos.
- Feierabend, T. and Eilks, I., (2010), Raising students' perception of the relevance of science teaching and promoting communication and evaluation capabilities using authentic and controversial socio-scientific issues in the Framework of climate change, *Sci. Educ. Int.*, **21**, 176–196.
- Field, A. P., (2024), *Discovering statistics using IBM SPSS statistics (6th edition)*, SAGE.
- Finn, J. D. and Zimmer, K. S., (2012), Student Engagement: What Is It? Why Does It Matter? in *Handbook of Research on Student Engagement*, ed. S. L. Christenson, A. L. Reschly and C. Wylie, Springer US, pp. 97–131.
- Fredricks, J. A., Blumenfeld, P. C. and Paris, A. H., (2004), School Engagement: Potential of the Concept, State of the Evidence, *Rev. Educ. Res.*, **74**(1), 59–109, DOI: [10.3102/00346543074001059](https://doi.org/10.3102/00346543074001059).
- Gibbs, K., (2023), Voices in practice: Challenges to implementing differentiated instruction by teachers and school leaders in an Australian mainstream secondary school, *Australian Educ. Res.*, **50**(4), 1217–1232, DOI: [10.1007/s13384-022-00551-2](https://doi.org/10.1007/s13384-022-00551-2).
- Güth, F. and van Vorst, H., (2023), Context-based Learning as a Method for Differentiated Instruction in Chemistry Education, in *Contributions from Science Education Research. FOSTERING SCIENTIFIC CITIZENSHIP IN AN UNCERTAIN WORLD: Selected papers*, ed. G. S. Carvalho, A. S. Afonso and Z. Anastácio, Springer International PU, vol. 13, pp. 153–169.
- Güth, F. and van Vorst, H., (2024), To choose or not to choose? Effects of choice in authentic context-based learning environments, *Eur. J. Psychol. Educ.*, 1–31, DOI: [10.1007/s12012-024-00798-6](https://doi.org/10.1007/s12012-024-00798-6).
- Habig, S., Blankenburg, J., van Vorst, H., Fechner, S., Parchmann, I. and Sumfleth, E., (2018), Context characteristics and their effects on students' situational interest in chemistry, *Int. J. Sci. Educ.*, **40**(10), 1154–1175, DOI: [10.1080/09500693.2018.1470349](https://doi.org/10.1080/09500693.2018.1470349).
- Hammond, J. and Gibbons, P., (2005), Putting scaffolding to work: The contribution of scaffolding in articulating ESL education, *Prospect*, **20**(1), 6–30, DOI: [10.3316/aeipt.143258](https://doi.org/10.3316/aeipt.143258).



- Hardy, I., Decristan, J. and Klieme, E., (2019), Adaptive teaching in research on learning and instruction. *Advance online publication*, DOI: [10.25656/01:18004](https://doi.org/10.25656/01:18004).
- Hattie, J. A. C., (2009), *Visible Learning: A synthesis of over 800 meta-analysis relating to achievement*, Routledge.
- Hofstein, A., Eilks, I. and Bybee, R., (2011), Societal Issues And Their Importance For Contemporary Science Education—A Pedagogical Justification And The State-Of-The-Art In Israel, Germany, And The USA, *In. J. Sci. Math. Educ.*, **9**(6), 1459–1483, DOI: [10.1007/s10763-010-9273-9](https://doi.org/10.1007/s10763-010-9273-9).
- Inkinen, J., Klager, C., Juuti, K., Schneider, B., Salmela-Aro, K., Krajcik, J. and Lavonen, J., (2020), High school students' situational engagement associated with scientific practices in designed science learning situations, *Sci. Educ.*, **104**(4), 667–692, DOI: [10.1002/sce.21570](https://doi.org/10.1002/sce.21570).
- IPCC, (2022), *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, H.-O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegria, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.), Cambridge University Press, Cambridge, UK and New York, NY, USA, p. 3056.
- Jang, H., Reeve, J. and Halusic, M., (2016), A New Autonomy-Supportive Way of Teaching That Increases Conceptual Learning: Teaching in Students' Preferred Ways, *J. Exp. Educ.*, **84**(4), 686–701, DOI: [10.1080/00220973.2015.1083522](https://doi.org/10.1080/00220973.2015.1083522).
- Janson, A., Söllner, M. and Leimeister, J. M., (2020), Ladders for Learning: Is Scaffolding the Key to Teaching Problem-Solving in Technology-Mediated Learning Contexts? *Acad. Manage. Learn. Educ.*, **19**(4), 439–468, DOI: [10.5465/amle.2018.0078](https://doi.org/10.5465/amle.2018.0078).
- Kim, N. J., Belland, B. R. and Walker, A. E., (2018), Effectiveness of Computer-Based Scaffolding in the Context of Problem-Based Learning for Stem Education: Bayesian Meta-analysis, *Educ. Psychol. Rev.*, **30**(2), 397–429, DOI: [10.1007/s10648-017-9419-1](https://doi.org/10.1007/s10648-017-9419-1).
- Klaver, L. T., van der Walma Molen, J. H., Sins, P. H. M. and Guérin, L. J. F., (2023), Students' engagement with Socio-scientific issues: Use of sources of knowledge and attitudes, *J. Res. Sci. Teach.*, **60**(5), 1125–1161, DOI: [10.1002/tea.21828](https://doi.org/10.1002/tea.21828).
- Klosterman, M. L. and Sadler, T. D., (2010), Multi-level Assessment of Scientific Content Knowledge Gains Associated with Socioscientific Issues-based Instruction, *Int. J. Sci. Educ.*, **32**(8), 1017–1043, DOI: [10.1080/09500690902894512](https://doi.org/10.1080/09500690902894512).
- Krapp, A., (2002), An educational-psychological theory of interest and its relation to SDT, in *The Handbook of Self-Determination Research*, pp. 405–427.
- Krapp, A. and Prenzel, M., (2011), Research on Interest in Science: Theories, methods, and findings, *Int. J. Sci. Educ.*, **33**(1), 27–50, DOI: [10.1080/09500693.2010.518645](https://doi.org/10.1080/09500693.2010.518645).
- Langelaan, B. N., Gaikhorst, L., Smets, W. and Oostdam, R. J., (2024), Differentiating instruction: Understanding the key elements for successful teacher preparation and development, *Teach. Teach. Educ.*, **140**, 104464, DOI: [10.1016/j.tate.2023.104464](https://doi.org/10.1016/j.tate.2023.104464).
- Lazega, E. and Snijders, T. A. B., (Eds.), (2016), *Methodos series: Vol. 12. Multilevel network analysis for the social sciences: Theory, methods and applications*, Springer.
- Letzel-Alt, V. and Pozas, M., (Eds.), (2023), *Differentiated Instruction Around the World: A Global Inclusive Insight*, Waxmann.
- Li, T., He, P. and Peng, L., (2023), Measuring high school student engagement in science learning: an adaptation and validation study, *Int. J. Sci. Educ.*, **46**(6), 524–547, DOI: [10.1080/09500693.2023.2248668](https://doi.org/10.1080/09500693.2023.2248668).
- Little, R. J. A. and Rubin, D. B., (2002), *Statistical Analysis with Missing Data*, Wiley, DOI: [10.1002/9781119013563](https://doi.org/10.1002/9781119013563).
- Luke, S. G., (2017), Evaluating significance in linear mixed-effects models in R, *Behav. Res. Methods*, **49**(4), 1494–1502, DOI: [10.3758/s13428-016-0809-y](https://doi.org/10.3758/s13428-016-0809-y).
- Maier, U. and Klotz, C., (2022), Personalized feedback in digital learning environments: Classification framework and literature review, *Comput. Educ.: Artif. Intell.*, **3**, 100080, DOI: [10.1016/j.caeai.2022.100080](https://doi.org/10.1016/j.caeai.2022.100080).
- Mathie, R. G., (2024), A Whole School Approach: A Synthesis of Interconnected Policy, Practice, and Research Conceptualisations, in *Whole School Approaches to Sustainability*, pp. 9–33, DOI: [10.1007/978-3-031-56172-6_2](https://doi.org/10.1007/978-3-031-56172-6_2).
- Meredith, W., (1993), Measurement Invariance, Factor Analysis and Factorial Invariance, *Psychometrika*, **58**(4), 525–543, DOI: [10.1007/BF02294825](https://doi.org/10.1007/BF02294825).
- Molenberghs, G. and Verbeke, G., (2025), Linear Mixed Models for Repeated Measures and Hierarchical Data, in *International Encyclopedia of Statistical Science*, Springer, Berlin, Heidelberg, pp. 1347–1356.
- MSB KLP, (2019), *Kernlehrplan für die Sekundarstufe I Gymnasium in Nordrhein-Westfalen. Chemie [Core curriculum for secondary level I grammar schools in North Rhine-Westphalia. Chemistry] (1st ed.)*, https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/198/g9_ch_klp_%203415_2019_06_23.pdf.
- Müller, T., Lichtinger, U. and Girg, R., (2015), *The Multi-GradeMultiLevel-Methodology and its Global Significance: Ladders of Learning - Scientific Horizons - Teacher Education. Theorie und Praxis der Schulpädagogik: Vol. 34, Prolog*, <https://edoc.ku.de/id/eprint/30129>.
- Mulvik, I., Prisbuisis, K., Siarova, H., Sabaliauskas, E. and Yemini, M., (2021), *Teachers have their say: Motivation, skills and opportunities to teach education for sustainable development and global citizenship*, https://www.researchgate.net/publication/359370642_Teachers_have_their_say_Motivation_skills_and_opportunities_to_teach_education_for_sustainable_development_and_global_citizenship.
- OECD, (2020), *Oecd Policy Responses to Coronavirus (COVID-19)*, *OECD Policy Responses to Coronavirus (COVID-19)*, Advance online publication, DOI: [10.1787/52b869f5-en](https://doi.org/10.1787/52b869f5-en).
- Olsson, D., Gericke, N. and Boeve-de Pauw, J., (2022), The effectiveness of education for sustainable development revisited – a longitudinal study on secondary students' action competence for sustainability, *Environ. Educ. Res.*, **28**(3), 405–429, DOI: [10.1080/13504622.2022.2033170](https://doi.org/10.1080/13504622.2022.2033170).
- Parchmann, I., Gräsel, C., Baer, A., Nentwig, P., Demuth, R. and Ralle, B., (2006), “Chemie im Kontext”: A symbiotic



- implementation of a context-based teaching and learning approach, *Int. J. Sci. Educ.*, **28**(9), 1041–1062, DOI: [10.1080/09500690600702512](https://doi.org/10.1080/09500690600702512).
- Pintrich, P. R., (2000), Chapter 14 - The Role of Goal Orientation in Self-Regulated Learning, in M. Boekaerts, P. R. Pintrich and M. Zeidner (Eds.), *Handbook of Self-Regulation*, Academic Press, pp. 451–502.
- Potvin, P. and Hasni, A., (2014), Interest, motivation and attitude towards science and technology at K-12 levels: a systematic review of 12 years of educational research, *Stud. Sci. Educ.*, **50**(1), 85–129.
- Pozas, M. and Schneider, C., (2019), Shedding Light on the Convoluted Terrain of Differentiated Instruction (DI): Proposal of a DI Taxonomy for the Heterogeneous Classroom, *Open Educ. Stud.*, **1**(1), 73–90, DOI: [10.1515/edu-2019-0005](https://doi.org/10.1515/edu-2019-0005).
- Pozas, M., Letzel, V. and Schneider, C., (2020), Teachers and differentiated instruction: exploring differentiation practices to address student diversity, *J. Res. Spec. Educ. Needs*, **20**(3), 217–230, DOI: [10.1111/1471-3802.12481](https://doi.org/10.1111/1471-3802.12481).
- Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D. and Soloway, E., (2004), A Scaffolding Design Framework for Software to Support Science Inquiry, *J. Learn. Sci.*, **13**(3), 337–386, DOI: [10.1207/s15327809jls1303_4](https://doi.org/10.1207/s15327809jls1303_4).
- Reeve, J., (2013), How students create motivationally supportive learning environments for themselves: The concept of agentic engagement, *J. Educ. Psychol.*, **105**(3), 579–595, DOI: [10.1037/a0032690](https://doi.org/10.1037/a0032690).
- Reeve, J. and Shin, S. H., (2020), How teachers can support students' agentic engagement, *Theory Pract.*, **59**(2), 150–161, DOI: [10.1080/00405841.2019.1702451](https://doi.org/10.1080/00405841.2019.1702451).
- Reeve, J. and Tseng, C.-M., (2011), Agency as a fourth aspect of students' engagement during learning activities, *Contemp. Educ. Psychol.*, **36**(4), 257–267, DOI: [10.1016/j.cedpsych.2011.05.002](https://doi.org/10.1016/j.cedpsych.2011.05.002).
- Reeve, J., Jang, H.-R., Shin, S. H., Ahn, J. S., Matos, L. and Gargurevich, R., (2022), When students show some initiative: Two experiments on the benefits of greater agentic engagement, *Learn. Instruct.*, **80**, 101564, DOI: [10.1016/j.learninstruc.2021.101564](https://doi.org/10.1016/j.learninstruc.2021.101564).
- Reinhold, F., Leuders, T., Loibl, K., Nückles, M., Beege, M. and Boelmann, J. M., (2024), Learning Mechanisms Explaining Learning With Digital Tools in Educational Settings: A Cognitive Process Framework, *Educ. Psychol. Rev.*, **36**(1), 1–21, DOI: [10.1007/s10648-024-09845-6](https://doi.org/10.1007/s10648-024-09845-6).
- Renninger, K. A. and Hidi, S. E., (2015), *The Power of Interest for Motivation and Engagement*, Routledge.
- Rienties, B., Giesbers, B., Tempelaar, D., Lygo-Baker, S., Segers, M. and Gijssels, W., (2012), The role of scaffolding and motivation in CSCL, *Comput. Educ.*, **59**(3), 893–906, DOI: [10.1016/j.compedu.2012.04.010](https://doi.org/10.1016/j.compedu.2012.04.010).
- Riess, W., Martin, M., Mischo, C., Kotthoff, H.-G. and Waltner, E.-M., (2022), How Can Education for Sustainable Development (ESD) Be Effectively Implemented in Teaching and Learning? An Analysis of Educational Science Recommendations of Methods and Procedures to Promote ESD Goals, *Sustainability*, **14**(7), 3708, DOI: [10.3390/su14073708](https://doi.org/10.3390/su14073708).
- Romine, W. L. and Sadler, T. D., (2016), Measuring Changes in Interest in Science and Technology at the College Level in Response to Two Instructional Interventions, *Res. Sci. Educ.*, **46**(3), 309–327, DOI: [10.1007/s11165-014-9452-8](https://doi.org/10.1007/s11165-014-9452-8).
- Rotgans, J. I. and Schmidt, H. G., (2018), How individual interest influences situational interest and how both are related to knowledge acquisition: A microanalytical investigation, *J. Educ. Res.*, **111**(5), 530–540, DOI: [10.1080/00220671.2017.1310710](https://doi.org/10.1080/00220671.2017.1310710).
- Ryan, R. M. and Deci, E. L., (2020), Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions, *Contemp. Educ. Psychol.*, **61**, 101860, DOI: [10.1016/j.cedpsych.2020.101860](https://doi.org/10.1016/j.cedpsych.2020.101860).
- Sadler, T. D., (Ed.), (2011), *Contemporary Trends and Issues in Sci. Educ.: Vol. 39. Socio-scientific Issues in the Classroom*, Springer Netherlands.
- Sadler, T. D., Romine, W. L. and Topçu, M. S., (2016), Learning science content through socio-scientific issues-based instruction: a multi-level assessment study, *Int. J. Sci. Educ.*, **38**(10), 1622–1635, DOI: [10.1080/09500693.2016.1204481](https://doi.org/10.1080/09500693.2016.1204481).
- Sadler, T. D. and Zeidler, D. L., (2005), Patterns of informal reasoning in the context of socioscientific decision making, *J. Res. Sci. Teach.*, **42**(1), 112–138, DOI: [10.1002/tea.20042](https://doi.org/10.1002/tea.20042).
- Saleh, M., Lazonder, A. W. and Jong, T. de (2005), Effects of within-class ability grouping on social interaction, achievement, and motivation, *Instruct. Sci.*, **33**(2), 105–119, DOI: [10.1007/s11251-004-6405-z](https://doi.org/10.1007/s11251-004-6405-z).
- Schnitzler, K., Holzberger, D. and Seidel, T., (2021), All better than being disengaged: Student engagement patterns and their relations to academic self-concept and achievement, *Eur. J. Psychol. Educ.*, **36**(3), 627–652, DOI: [10.1007/s10212-020-00500-6](https://doi.org/10.1007/s10212-020-00500-6).
- Schnotz, W. and Rasch, T., (2005), Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning, *Educ. Technol. Res. Dev.*, **53**(3), 47–58, DOI: [10.1007/BF02504797](https://doi.org/10.1007/BF02504797).
- Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G. and Friedman, T., (2018), *Becoming Citizens in a Changing World*.
- Shao, J., Chen, Y., Wei, X., Li, X. and Li, Y., (2023), Effects of regulated learning scaffolding on regulation strategies and academic performance: A meta-analysis, *Front. Psychol.*, **14**, 1110086, DOI: [10.3389/fpsyg.2023.1110086](https://doi.org/10.3389/fpsyg.2023.1110086).
- Sinatra, G. M., Heddy, B. C. and Lombardi, D., (2015), The Challenges of Defining and Measuring Student Engagement in Science, *Educ. Psychol.*, **50**(1), 1–13, DOI: [10.1080/00461520.2014.1002924](https://doi.org/10.1080/00461520.2014.1002924).
- Smale-Jacobse, A. E., Meijer, A., Helms-Lorenz, M. and Maulana, R., (2019), Differentiated Instruction in Secondary Education: A Systematic Review of Research Evidence, *Front. Psychol.*, **10**, 2366, DOI: [10.3389/fpsyg.2019.02366](https://doi.org/10.3389/fpsyg.2019.02366).
- Stuckey, M., Hofstein, A., Mamlok-Naaman, R. and Eilks, I., (2013), The meaning of 'relevance' in science education and its implications for the science curriculum, *Stud. Sci. Educ.*, **49**(1), 1–34, DOI: [10.1080/03057267.2013.802463](https://doi.org/10.1080/03057267.2013.802463).



- Sun, Y. and Xiao, L., (2021), Research trends and hotspots of differentiated instruction over the past two decades (2000–2020): a bibliometric analysis, *Educ. Stud.*, **50**(2), 186–202, DOI: [10.1080/03055698.2021.1937945](https://doi.org/10.1080/03055698.2021.1937945).
- Tahtinen, L. and United Nations Department of Economic and Social Affairs, (2020), *SDG Good Practices: a Compilation of Success Stories and Lessons Learned in SDG implementation (First Edition)*, United Nations Department of Economic and Social Affairs (UN-DESA), <https://wfp.tind.io/record/68306>.
- Tekin, N., Aslan, O. and Yilmaz, S., (2016), Research Trends on Socioscientific Issues: A Content Analysis of Publications in Selected Science Education Journals, *J. Educ. Train. Stud.*, **4**(9), DOI: [10.11114/jets.v4i9.1572](https://doi.org/10.11114/jets.v4i9.1572).
- Tijmstra, J. and Bolsinova, M., (2023), The Hierarchical Model for Response Times: Advantages, Limitations, and Risks of Its Use in Measurement Practice, *Essays Contemp. Psychomet.*, 307–325, DOI: [10.1007/978-3-031-10370-4_16](https://doi.org/10.1007/978-3-031-10370-4_16).
- Tomlinson, C. A., Brighton, C. M., Hertberg, H., Callahan, C., Moon, T., Brimijoin, K., Conover, L. A. and Reynolds, T., (2003), Differentiating Instruction in Response to Student Readiness, Interest, and Learning Profile in Academically Diverse Classrooms: A Review of Literature, *J. Educ. Gifted*, **27**(2-3), 119–145, DOI: [10.1177/016235320302700203](https://doi.org/10.1177/016235320302700203).
- Tuckman, B. W., (2007), The effect of motivational scaffolding on procrastinators' distance learning outcomes, *Comput. Educ.*, **49**(2), 414–422, DOI: [10.1016/j.compedu.2005.10.002](https://doi.org/10.1016/j.compedu.2005.10.002).
- United Nations Department of Economic and Social Affairs, (2025), *The Sustainable Development Goals Report 2025*.
- UNESCO, (2017), *Education for sustainable development goals: learning objectives*, UNESCO.
- van de Pol, J., Volman, M. and Beishuizen, J., (2010), Scaffolding in Teacher–Student Interaction: A Decade of Research, *Educ. Psychol. Rev.*, **22**(3), 271–296, DOI: [10.1007/s10648-010-9127-6](https://doi.org/10.1007/s10648-010-9127-6).
- van de Pol, J., Volman, M., Oort, F. and Beishuizen, J., (2015), The effects of scaffolding in the classroom: support contingency and student independent working time in relation to student achievement, task effort and appreciation of support, *Instruct. Sci.*, **43**(5), 615–641, DOI: [10.1007/s11251-015-9351-z](https://doi.org/10.1007/s11251-015-9351-z).
- van Vorst, H., (2018), Structuring learning processes by ladders of learning: results from an implementation study, *Chem. Educ. Res. Pract.*, **19**(4), 1081–1095, DOI: [10.1039/C8RP00078F](https://doi.org/10.1039/C8RP00078F).
- van Vorst, H. and Aydogmus, H., (2021), One context fits all? – analysing students' context choice and their reasons for choosing a context-based task in chemistry education, *Int. J. Sci. Educ.*, **43**(8), 1250–1272, DOI: [10.1080/09500693.2021.1908640](https://doi.org/10.1080/09500693.2021.1908640).
- van Vorst, H., Dorschu, A., Fechner, S., Kauertz, A., Krabbe, H. and Sumfleth, E., (2015), A theoretical framework for categorizing context-based tasks in science education, *Z. Didaktik Naturwissenschaften*, **21**(1), 29–39, DOI: [10.1007/s40573-014-0021-5](https://doi.org/10.1007/s40573-014-0021-5).
- Waltner, E.-M., Scharenberg, K., Hörsch, C. and Rieß, W., (2020), What Teachers Think and Know about Education for Sustainable Development and How They Implement it in Class, *Sustainability*, **12**(4), 1690, DOI: [10.3390/su12041690](https://doi.org/10.3390/su12041690).
- Wigfield, A. and Eccles, J. S., (2000), Expectancy–Value Theory of Achievement Motivation, *Contemp. Educ. Psychol.*, **25**(1), 68–81, DOI: [10.1006/ceps.1999.1015](https://doi.org/10.1006/ceps.1999.1015).
- Wood, D., Bruner, J. S. and Ross, G., (1976), The role of tutoring in problem solving, *J. Child Psychol. Psychiatr., Allied Discipl.*, **17**(2), 89–100, DOI: [10.1111/j.1469-7610.1976.tb00381.x](https://doi.org/10.1111/j.1469-7610.1976.tb00381.x).
- Zimmerman, B., (2008), Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects, *Am. Educ. Res. J.*, **45**, 166–183, DOI: [10.3102/0002831207312909](https://doi.org/10.3102/0002831207312909).
- Zulyusri, Z., Indah, A. and Santosa, T. A., (2022), Meta-analysis: The Effectiveness of Using Socio-scientific Issues on Science Literacy and Students' Higher-Order Thinking Ability in Science Learning, *Literacy: Int. Sci. J. Soc., Educ., Hum.*, **1**(2), 94–105, DOI: [10.56910/literacy.v1i2.221](https://doi.org/10.56910/literacy.v1i2.221).

