Discover Education

Research

The effectiveness of the teaching program "scientific creativity in practice"

Kurt Haim[1] · **Wolfgang Aschauer**[1] · **Christoph Weber**[1]

Received: 23 July 2024 / Accepted: 22 November 2024

Abstract

This study examines the impact of the Scientific Creativity in Practice program on students' scientific creativity. This comprehensive teaching program aims to foster general creative skills, domain-specific competencies, and certain personality traits in science education. The program includes a variety of interventions and reflection tools that focus on divergent thinking, problem-solving, bisociation, imagination, and metacognition, which promotes strategies for developing, selecting, and evaluating creative solutions to scientific problems. The study used a two-group, repeated-measures design in Austrian secondary schools, with 26 classes taught using teaching techniques from the program and 12 control classes using traditional teaching methods. The effectiveness of the teaching program was assessed using the Divergent Problem-Solving Ability test, which was administered at the beginning and end of the school year. The results showed that the intervention classes achieved significantly higher scores, indicating improved divergent problem-solving ability.The study also emphasized the importance of intervention fidelity, with the quality of implementation (adherence) having a significantly higher impact on outcomes than the quantity (number of interventions). The results confirm the effectiveness of the Scientific Creativity in Practice program in fostering academic creativity and emphasize the importance of a well-structured teaching program. Future research will focus on measuring the impact of individual techniques of the program on different aspects of scientific creativity.

Keywords Scientific creativity · Divergent thinking · Intervention study · Intervention fidelity

1 Introduction

Global challenges like climate change and environmental degradation require a change in the way we think and act [1]. In order to shape our world in an innovative, resource-conserving and sustainable way in the future, young people need to learn at school how to solve problems for which there are no ready-made strategies [2, 3].

Fostering scientific creativity in science subjects is, therefore, crucial for preparing students to face the challenges of the twenty-first century. Creative skills enable students to develop innovative and sustainable solutions for complex problems [1, 4].

Scientific creativity encompasses both general creative abilities and domain-specific competencies, as well as typical personality traits [5–7]. General creative abilities include divergent thinking, bisociative thinking, and analogical

---

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s44217-024-00368-4.

✉ Kurt Haim, kurt.haim@ph-ooe.at | [1]Department of Science Education, University of Education Upper Austria, Linz, Austria.

thinking, which are crucial for problem-solving and innovation [8, 9]. Domain-specific competencies involve the ability to define scientific problems, formulate hypotheses, and conduct experiments [10]. Moreover, programs aimed at fostering scientific creativity also promote personality traits such as openness to new ideas, resilience, and the courage to break new ground [11]. Studies have shown that students who participate in such programs not only achieve higher creative performance but are also better prepared to handle unpredictable and complex situations [12, 13].

Understanding and fostering these skills comprehensively is essential to enable students to succeed in an ever-changing world and to drive scientific innovation [14]. Curricula and teaching methods must be adapted to integrate creative thinking strategies in order to strengthen scientific creativity in science subjects [6].

In recent years, several training programs have been developed and evaluated to promote creativity in the natural sciences. Examples include the Creative-Inquiry-Based Science Teaching program [15] and the Learn to Think program [12]. Many of these deal with only one or two aspects of scientific creativity and only one type of action. Usually, these programs focus on inquiry-based learning, problem-based learning or collaborative approaches.

In contrast to other programs, Haim and Aschauer [16] developed a teaching program for science subjects called Scientific Creativity in Practice (SCIP; previously known as flex-based learning). This program includes a comprehensive collection of tools to promote the most important aspects of scientific creativity.

A unique feature of the SCIP program is that it uses specially developed practice techniques that link general creative skills with domain-specific skills. In this way, divergent thinking techniques are used to build and consolidate scientific domain knowledge. While the majority of the exercises are designed to promote creative thinking, the SCIP program also includes tools with problem-based tasks that must be solved experimentally. At the same time, the program consists of numerous reflection tools to highlight and strengthen important creative personality traits. Therefore, the SCIP program is not only aimed at promoting a few aspects of scientific creativity but is a comprehensive training program that helps students to develop scientific creativity in the best possible way.

In addition to providing an introduction to scientific creativity and a brief description of known support approaches, particularly the SCIP program, this paper reports the results of an intervention study, which was guided by the following two research questions:

(1) How does the SCIP program influence students' scientific creativity, specifically their divergent problem-solving abilities in science subjects?
(2) How does intervention fidelity, especially adherence and dosage, influence students' outcomes related to divergent problem-solving abilities?

2 Theoretical framework

2.1 Scientific creativity

Following the standard definition of creativity [17], scientific creativity is the ability to develop novel products and ideas with scientific value [5, 18]. Any thought or behavior that leads to new and useful products in science can be called scientific creativity. Thus, new theories and hypotheses, original research designs, innovative procedures etc. can be characterized as creative products in science [19].

It should be mentioned that creative products can reach different levels of creative magnitude, for which the terms big-c (outstanding contributions of experts), little-c (everyday products) and mini-c (construction of personal knowledge and understanding) have been introduced [20, 21]. In each case, 'c' stands for creativity. In this sense, creative learning processes or creative products that have emerged in practice-oriented lessons at school can be classified as mini-c and little-c [22]. Even if other young people have previously generated the same processes or products, they are new to them personally, which satisfies the definition of creativity.

There is a consensus in creativity research that domain-specific competencies, general creative abilities as well as motivation and personality traits (e.g., the willingness to take risks; [23]) are responsible for creative performance in science [5–7].

As mentioned in the introduction, examples of domain-specific competencies include the ability to identify problems [8, 10], problem-solving abilities [7], and the ability to formulate and test hypotheses [5, 10].

In order to develop these domain-specific skills, certain general creative abilities are required. These include divergent thinking, the ability to make original associations and bisociations, understanding and working with analogies, imagination and metacognitive abilities. These five skills are described in more detail below.

- Divergent thinking (DT) is an essential predictor of creative problem-solving ability [5, 8, 24]. DT ability is measured by fluency (number of responses given), flexibility (assigning ideas to different categories) and originality (uniqueness of an idea) [25]. The use of DT allows an individual to generate a variety of solutions to a problem from a wide range of perspectives, increasing the likelihood of successful, creative problem-solving [26].
- Original associations and bisociations are links between two aspects within a theme as well as links between two very different topics [27]. Both can be regarded as elementary cognitive processes for creativity [9, 28], as creative ideas arise from the fertile association of unrelated concepts, themes or images [29].
- Through analogies, different concepts are compared and similarities between them are discovered. Analogies can be very helpful in problem-solving when similarities are discovered between different problems and a solution strategy is known for at least one of the problems [30]. Thus, creating and interpreting analogies can significantly increase an individual's scientific creativity [2, 6].
- Imagination can be used to mentally remove oneself from the current time, location and circumstances, to develop internal images or scenarios and to create fictional worlds [31]. Since creativity requires the interaction between logical thinking and imagination, imaginative ability is one of the central elements of scientific creativity [2, 32].
- Metacognitive skills include cognitive knowledge, such as knowledge about creative thinking styles [9, 33, 34], and cognitive regulation, such as assessing one's strengths and weaknesses during creative teamwork [33, 35]. Creative thinking processes are strongly dependent on a high level of metacognition, as they can only succeed through comprehensive planning, management and control of cognitive processes [36, 37].

2.2 Programs to foster scientific creativity

Research on creativity has been ongoing since the 1950s [17], and research approaches related to educational support programs have had the greatest and most lasting social impact [38]. This includes programs that have focused on fostering creativity in students in science classrooms [2].

A meta-analysis by Bi et al. [39] examined the effectiveness of 17 intervention studies seeking to promote scientific creativity conducted between 1992 and 2019. In this meta-analysis, four types of interventions were found to be effective: problem-solving, collaborative learning, conceptual construction and scientific reasoning.

The problem-solving group includes interventions for promoting students' problem-solving skills. This involves a range of cognitive processes, from problem identification to creative idea generation and problem-solving. An example of this is an intervention called Scientific Process Skills [40], which promotes skills like problem definition, observation and analysis, all of which are important for solving scientific problems.

Collaborative learning focus on promoting knowledge sharing in group discussions, such as sharing one's thoughts, listening to other group members and constructing knowledge creatively to achieve common goals. For example, Siew and Chin [41] integrated problem-based learning and collaborative learning to promote scientific creativity.

The conceptual construction group includes interventions to help students develop a knowledge structure for scientific concepts. The associated accumulation of knowledge is an essential component of scientific creativity, especially because it can lead to a greater number of possible combinations of scientific knowledge [18, 42]. For example, Nasiri [43] used computer-assisted instruction to teach basic concepts of three physics topics—heat, electrostatics and electric current—and examined the effects on changes in scientific creativity.

The scientific reasoning group contains thinking-training activities that are hypothetic-deductive in nature. These activities include observing phenomena, developing plausible explanations, deriving conclusions and then designing and conducting experiments to confirm, reject or revise the hypotheses [44]. Among these interventions is the Learn to Think program [12], which aims to increase scientific creativity by fostering relevant creative thinking strategies, such as working with analogies. Another example is the Divergent Thinking Training program created by Sun et al. [45], which is designed to help students master divergent thinking strategies and use them in creative scientific work.

The results of the meta-analysis conducted by Bi et al. [39] showed that interventions targeting problem-solving and scientific reasoning exhibited the most significant effects on scientific creativity.

A further systematic review of 30 studies by Sidek et al. [46] found that pedagogical strategies such as teaching creative thinking techniques, problem-based, project-based, model-based, ICT-based (Information and Communication

Technology), integrated STEM-based (Science, Technology, Engineering and Math), and collaborative learning were found to improve scientific creativity among students. Creative thinking skills, as listed by Hu et al. [12], include analogy, reorganization, brainstorming, breaking the set and transference. A learning experience that is based on problems, projects, and modelling enables students to construct their own knowledge, allowing them to express their diversity including scientific creativity. In some studies, the use of ICT educational media promoted students' creativity, as access to up-to-date data and a wealth of knowledge enabled them to brainstorm ideas more easily. Science education also seems to promote scientific creativity. It is an interdisciplinary approach that combines science, technology, engineering and mathematics. Through this approach students are motivated to make connections between subject matter and the real world. The integration of different disciplines in science education and the high degree of active learning improves research skills, problem-solving skills and, thus, scientific creativity [47]. Collaborative learning can also encourage students' creativity, as pupils can complement each other in generating ideas and, thus, address a wide range of perspectives. In summary, the approaches discussed have similar characteristics, such as a strong student-centeredness where students have more autonomy in their learning process. In addition, brainstorming and reasoning skills are usually included in problem- or project-based interventions. Brainstorming and reasoning are thinking techniques that are very useful in developing creativity.

However, as can be seen from the examples above, most training programs include only one or two aspects of scientific creativity. Furthermore, the fact that students are given the opportunity to brainstorm does not guarantee that they will be able to change perspectives and come up with original ideas. For this reason, the SCIP program was selected for this study, as it is based on a more comprehensive approach. It contains a collection of many different tools to promote the flexibility and originality of solution ideas and to strengthen creative personality traits.

2.3 Scientific creativity in practice (SCIP)

SCIP is a new and innovative approach to science teaching that focuses on the comprehensive promotion of scientific creativity (previously known as flex-based learning; [16]). The novel aspect of the SCIP program lies in the simultaneous promotion of general creative thinking processes, science-specific skills, metacognition and strengthens creative personality traits through specially developed teaching tools. The SCIP program includes about 10 tools to promote strategies for developing, evaluating and selecting creative solutions to scientific problems. The selection of the SCIP tool for the program is closely linked to the content of common curricula.

The special feature of the SCIP tools is that each one promotes a different aspect of general creativity and at the same time builds up scientific knowledge. While the so-called Thinkflex tools specifically promote divergent thinking, the Woseco tool trains associative thinking. With Live Acts, the aim is to improve imagination and fantasy when dealing with scientific problems. Flexperiments are used to strengthen divergent thinking skills in scientific experiments. It shows students how to plan and carry out open-ended experiments and critically reflect on their results. Special reflection tools such as 'Shorty & Flexy' and 'Role models' are used to improve students' metacognitive awareness. Finally, the SCIP program also includes reflection tools to strengthen personality traits such as fault tolerance and perseverance, which play an important role in creative processes. What makes the SCIP program outstanding compared to other programs is the great diversity of tools used to develop the most important factors of creative thinking and acting.

The SCIP tools are designed in such a way that an interplay between general creative skills and scientific skills is required to solve the tasks. In this way, general creative skills are used to consolidate scientific knowledge. Furthermore, general creative skills are used to reinforce specific scientific methodologies such as questioning, hypothesizing and problem-solving, which are essential for creative performance.

The SCIP tools used in this intervention study are briefly described below.

THINKFLEX includes cognitive thinking tasks designed to increase divergent thinking and, in particular, mental flexibility. A so-called perspective check is used as a support tool, which guides students from one thinking perspective to another and facilitates thinking in different categories. Many THINKFLEX tasks are formulated to prompt students to practice scientific thinking by forming hypotheses, finding arguments and reasons or drawing conclusions. Furthermore, THINKFLEX tasks also promote collaborative learning, as the work steps follow the listen–think–pair–share setting [48].

WOSECO is an acronym for word-sentence-constructions. This tool challenges students to combine technical terms from different chapters of a subject into correct sentences. Firstly, this tool helps to consolidate subject knowledge and linguistic competence, and secondly, it trains the ability to form original associations [49].

LIVE ACTS are a type of performance in which students are asked to represent a scientific phenomenon or a molecular process at particle level. To do this, the students represent the nanoworld with their bodies and given utensils, such as strings, bags or balls. By alternating between the perspectives of the macrocosm and the microcosm, the students' imagination and analogical thinking are enhanced. After the demonstration, a "what if" question is added, and the students are asked to formulate a hypothesis and justify it using this tool. LIVE ACTS not only promote students' imagination and analogical thinking but also emphasize the high value of these skills in formulating scientific hypotheses. As LIVE ACTS require a high degree of interaction between students within a group, this instrument also makes an important contribution to collaborative learning.

FLEXPERIMENTS are a special form of problem-solving experiments that involve realistic problems. Of all the SCIP tools, FLEXPERIMENTS challenge a student's scientific creativity in the most authentic way, as they require all aspects of scientific creativity. From problem identification, creative brainstorming, decision-making, independent planning and implementation of an experiment to critical reflection on the results, students experience all the important phases of a creative process. The particular challenge in a FLEXPERIMENT is not to solve a problem in just one way with the help of selected materials but to find as many different solutions as possible and implement them experimentally in the group. By finding multiple solutions to a problem, students engage in both divergent and critical thinking, as each solution variant must also be evaluated. Thus, according to Bi et al. [39], this tool trains not only creative problem-solving but also divergent thinking and collaborative working.

2.4 Research aims

The study described here had two aims:

Firstly, to investigate the extent to which SCIP increases students' scientific creativity, specifically their divergent problem-solving abilities in science.

Secondly, to investigate the influence of intervention fidelity, especially adherence and degree, on students' outcomes related to divergent problem-solving abilities in science.

3 Methodology

3.1 Procedures

Before the study began, teachers were informed about the procedure and objectives of the research project and asked whether they would like to participate with their pupils. There were no exclusion criteria.

For an intervention to be implemented with fidelity, adequate training and supervision by interventionists during the intervention study are required [50–52]. Therefore, all participating teachers participated in a one-year in-service teacher-training program regarding SCIP as a requirement for the study. The teacher training program was structured into alternating input, implementation, and reflection phases. In the input phases, both the theoretical concepts of scientific creativity were discussed and the individual SCIP techniques were introduced. These theoretical and practical inputs were provided in September 2018 and March 2019, each in a two-day session. In the implementation phases, the teachers applied the.

SCIP techniques in their own lessons. The teachers reflected on their experiences in joint meetings held in November, February, April and June. For the study, the participating science teachers were provided with prepared worksheets for the subjects of chemistry, physics and biology for each SCIP tool (for concrete examples of what the worksheets looked like, see Haim and Aschauer [53]).

All procedures performed in this study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the institutional review board of the Upper Austrian Department of Education.

3.2 Participants

The study was conducted using a two-group repeated-measures design. In 26 intervention classes (570 students, 54% males) in Austrian secondary schools (24% 7th graders and 76% 8th graders), the teachers continuously implemented the SCIP tools described above throughout the 2018–19 school year. In contrast, no specific interventions promoting scientific creativity were implemented in 12 control classes (CTRL; 223 students, 57% males, 50% grade 7 and 50% grade

8). Half of these classes were taught by the same teachers. The Divergent Problem-Solving Ability in Science test (DPAS; [53]) was administered online both at the beginning (t1, September 2018) and end of the school year (t2, June 2019).

Due to the procedure, the students are a convenience sample. Informed consent was obtained from all individual students and their parents.

3.3 Measures

3.3.1 Divergent problem-solving ability in science

The DPAS was used as the outcome variable, as this test was developed specifically to assess divergent problem-solving ability in science in intervention studies. The DPAS consists of two subtests: SUBTEST A (SA; 6 items) and SUBTEST B (SB; 7 items). SA requires students to name, for example, reasons for failed processes or the consequences of scientific phenomena. SB requires students to find solutions for experimental tasks, with only a limited number of materials. A validation study [53], in which 1106 secondary school students from Austria participated (the 793 students included in the current study are a subsample of the validation sample), has provided extensive support for the validity and reliability ($\omega_{SA\_t1} = 0.79$, $\omega_{SA\_t2} = 0.83$, $\omega_{SB\_t1} = 0.69$, and $\omega_{SB\_t2} = 0.78$) of the DPAS. In addition, the results supported longitudinal and group-specific (intervention vs. control group) factorial invariance. Overall, the validation study confirmed that the DPAS is capable of validly measuring divergent problem-solving ability in a two-group repeated-measures design.

Notably, not all DPAS items were used at each time point (4 SB and SA items at t1 and 5 SB and SA items at t2). To ensure a common metric for the DPAS scores, linking items (3 for SA and 2 for SB) were used.

The students were instructed to find as many solutions as possible; however, due to feasibility issues, there was a time limit established per item. For each item, the creativity quotient (CQ; [54]) was calculated. CQ is a creativity measure that encompasses both fluency (i.e., the number of appropriate responses) and flexibility (i.e., the number of different categories into which the answers or ideas fall). It is calculated as follows:

$$CQ \ {=}^{2} \log \left[\left(1+n_1\right)\left(1+n_2\right)\ldots\left(1+n_C\right) \right]$$

C is the number of different categories for an item and n_i is the number of responses within category i. Due to the calculation formula, CQ gives greater weight to responses that belong to distinct categories and thus more weight on flexibility.

For each item all responses were rated by two raters, who are members of the research team, based on the scoring guide of the validation study [53]. The CQ scores are averaged over raters and were used for all analyses.

For the interrater reliability, the intraclass correlation (ICC) for absolute agreement based on a two-way mixed model was estimated. As in the validation study [53], the interrater reliability achieved good values between ICC = 0.873 and ICC = 0.992.

3.3.2 Intervention fidelity

In various disciplines, from health to science education, intervention programs are not always implemented by practitioners as originally planned. Differences between the intended and implemented intervention, in other words how well an intervention is implemented compared to the original program design, can be described in terms of integrity or fidelity [51, 55]. Procedures for operationalizing intervention fidelity commonly focus on how much and how well a program is implemented [51].

Dosage, which refers to quantity, number, length and frequency, was measured by the total number of individual SCIP tools that were utilized based on teachers' self-reported surveys.

Regarding quality, our focus was on adherence, the degree to how well the interventions were delivered based on the evidence-based characteristics of the SCIP tools. As in other studies (e.g., [39, 56, 57]), adherence was determined using checklist-based scores, rated independently by two raters. Based on teachers' self-reported surveys, portfolios and feedback in the reflection sessions, two members of the research team rated the degree of adherence for each teacher. For this purpose, a checklist was used that covered the following points:

(1) The specified minimum number of interventions for each SCIP tool was fulfilled; (2) All phases were carried out as intended; (3) The prerequisites for the intervention were satisfied; (4) Students' self-reflected on their divergent thinking skills; and (5) The predetermined tasks and worksheets were used.

For each individual intervention, the critical components were scored from 0 = "not implemented as intended" to 2 = "completely implemented as intended." Regarding the degree of adherence, the scores of all interventions were added up and divided by the maximum achievable score, resulting in a percentage scale ranging from 0 to 100%. The interrater reliability—assessed by the intraclass correlation (ICC) for absolute agreement based on a two-way mixed model—was high (0.907). In the case of divergent values, both raters reached a consensus.

The second aspect of quality is program differentiation. Differentiation ensures that only the specific program activities are the cause of the observed changes [58]. Based on the teachers' self-reports, control classes were rated from 1 = "several program-like tools were used" to 3 = "no program-like tools were used at all."

The quality of delivery and participant responsiveness are usually assessed through lesson observations. Due to personal and time resources, these two elements were not surveyed in this study. And there is also the possibility that the presence of researchers in the classroom influences the behavior of teachers and students [59].

3.3.3 Control variables

In all analyses, we controlled for grade level, gender and grade point average (GPA) in physics and mathematics, as these variables are likely correlated with scientific creativity or science achievement more generally [53, 60].

3.4 Analyses

As the SCIP interventions were administered at the class level, a multilevel modeling approach using Mplus 8.1 [61] was applied in this study. Moreover, the outcome variable (divergent problem-solving ability) corresponds to a latent construct, supporting the use of structural equation modeling (SEM; e.g., [62]), which allows for the inclusion of both measurement and structural models (e.g., SCIP effects).

However, multilevel SEM requires a large sample size at the upper level (level 2 units, i.e. classroom level in this study). It is recommended that there should be at least 50 upper level units, as smaller numbers of level 2 units result in unacceptable high rates of convergence problems and estimation errors ([63, 64]). Thus, given the relatively small sample size at the classroom level (n = 38) in this study, we refrained from simultaneously considering the measurement model using CQ-scores of the DPAS items as indicators of the outcome variable and the SCIP effects. Instead, we used a two-step approach. First, we fit longitudinal factor models for both DPAS subscales assuming strong longitudinal factorial invariance, thus assuring the same metric for the DPAS scores at t1 and t2 (for details on model specification, model fit and testing longitudinal invariance see Figure S1 and Table S1 in the Online Resource). Based on the longitudinal factor model we generated 20 plausible values (PVs) to be used as outcome variables in subsequent analyses. PVs—the Bayesian equivalent of factor scores—are multiple imputations for latent variables and are also used in large-scale studies (e.g., TIMSS or PISA; [66]). PVs are considered superior to factor scores, as they, for example, do not underestimate the (co)variance of latent variables [65] and yield similar results (e.g. regression coefficients) to SEM [66].

Secondly, we used standard multilevel modeling to answer the research questions. Analyses were conducted separately for each of the two DPAS subtests. To investigate the effects of SCIP, group membership (intervention vs. control group) was used as a class-level (L2) predictor of students' outcomes at t2. At the student level (L1), the respective DPAS score (i.e., SB or SA) at t1 and the control variables were used as predictors. Grade level was used at L1 due to the inclusion of some cross-grade classes. To investigate the effects of intervention fidelity, we focused only on the SCIP classes. We used the same L1 predictors but used the intervention fidelity measures—concept adherence and number of performed SCIP tools—as L2 predictors. Due to the small relative sample size for this investigation (n = 26), analyses were conducted separately for both fidelity measures. Notably, we z-standardized the DPAS scores and all other continuous variables. Thus, the reported effects are in a standardized (i.e., effect size) metric. The SCIP effects were the mean differences between the SCIP and CTRL classes (comparable to Cohens d), standardized in relation to the total outcome variance (L1 and L2), while the fidelity effects were standardized in relation to the L2 fidelity variance and the total outcome variance.

4 Results

4.1 Differences between SCIP and the control classes at T1

Table 1 presents the descriptive results for the SCIP and CTRL groups at t1. Notably, there were small but insignificant differences between the SCIP and CTRL groups in the DPAS subtests ($d_{SB} = 0.308$, $p > 0.05$; $d_{SA} = 0.218$, $p > 0.05$). Moreover, there were no gender or GPA differences. However, the SCIP and CTRL groups differed in terms of grade level. While 76% of the SCIP students attended 8th grade, only half of the CTRL students were in the 8th grade. However, this effect was also not significant (Phi = 0.248, $p > 0.05$).

4.2 Intervention Fidelity

On average, 9.38 (SD = 3.46) interventions were carried out in the school year. In line with the relatively high SD, the number of interventions ranged from four to 18. The degree of adherence to the SCIP program was high at 91% (SD = 7.00). Regarding program differentiation, there was only one control class in which a teacher implemented an SCIP intervention and another in which a teacher implemented multiple SCIP interventions. All other teachers were compliant and did not implement any SCIP-like interventions. Therefore, program differentiation, which was evaluated for each teacher of the control group on a three-point scale ranging from 1 = "serval program-like tools were used" to 3 = "no program-like tools were used at all," achieved an average value of 2.75 (SD = 0.62). Thus, the conditions for the control classes were very well fulfilled by the teachers.

4.3 Effects of SCIP on student outcomes

The results regarding the effects of SCIP are reported in Table 2. As shown in Table S2, for both DPAS subtests, there was considerable variation between classes at t2 ($ICC_{SB} = 0.214$, $p < 0.001$; $ICC_{SA} = 0.196$, $p < 0.001$). However, this between-class variation could largely be attributed to T1 differences—supporting the use of multilevel modeling—and significant variations between classes after controlling for pretest differences of the respective subtests ($ICC_{SB} = 0.043$, $p < 0.05$; $ICC_{SA} = 0.070$, $p < 0.01$; see Table S2). Further, for both subtests, we found significant effects of the SCIP intervention ($b_{SB} = 0.223$, $p < 0.05$; $b_{SA} = 0.305$, $p < 0.001$), accounting for 29% (SB) and 40% (SA) of the between-class variance. Thus, controlling for students' baseline scores in the respective subtests (and the sets of control variables), students in the SCIP classes scored higher on both DPAS subtests at t2 than students in the control classes. As noted above, two control classes deviated from the CTRL conditions. Thus, we conducted further analyses excluding these classes. Again, the SCIP effects were significant and, as expected, somewhat higher than the effects considering all CTRL classes ($b_{SB} = 0.339$, $p < 0.01$, $R^2 = 0.479$; $b_{SA} = 0.401$, $p < 0.001$, $R^2 = 0.482$).

Table 1 Descriptive statistics of the study variables at time 1 (t1)

	SCIP	CTRL	Difference
	M (SD)/%	M (SD)/%	Effect size[a] [95%-CI]
DPAS-SB_{t1}	1.077 (0.479)	0.934 (0.440)	0.305 [– 0.039, 0.649]
DPAS-SA_{t1}	3.108 (1.117)	2.866 (1.085)	0.217 [– 0.077, 0.510]
GPA_MS	2.472 (1.000)	2.483 (0.981)	– 0.011 [– 0.240, 0.217]
Gender (Male)	54.4%	57.0%	0.023 [– 0.117, 0.071]
Grade level			
7th	24.2%	49.8%	0.248 [– 0.083, 0.578]
8th	75.8%	50.2%	
N (students)	570	223	
N (classes)	26	12	

[a] Cohens d for continuous variables and Phi for categorical variables. CI = confidence interval. *Type = Complex* command in M*plus* was used to adjust confidence intervals for the clustered data (i.e., students nested in classes)

Table 2 Effects of SCIP on DPAS subtests

	SB	SA
	b (SE) [95%-CI]	b (SE) [95%-CI]
Within Class		
Pretest	0.618*** (0.044) [0.532, 0.704]	0.546*** (0.041) [0.466, 0.626]
GPA_MS	– 0.112** (0.040) [– 0.191, – 0.033]	– 0.221*** (0.042) [– 0.303, – 0.139]
Gender (Male)	– 0.186* (0.023) [– 0.347, –0.026]	– 0.318*** (0.071) [– 0.458, – 0.179]
Grade	0.040 (0.112) [– 0.179, 0.260]	0.187 (0.097) [– 0.002, 0.377]
Residual Variance	0.441*** (0.037) [0.368, 0.514]	0.415*** (0.036) [0.345, 0.484]
Between Class		
SCIP	0.223* (0.113) [0.002, 0.443]	0.305*** (0.091) [0.128, 0.483]
Intercept	– 0.161 (0.087) [– 0.331, 0.009]	– 0.219*** (0.068) [– 0.352, – 0.086]
Residual Variance	0.028* (0.013) [0.002, 0.053]	0.031** (0.012) [0.007, 0.055]
R^2 Within	0.515	0.537
R^2 Between	0.290	0.401

CI = confidence interval. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

4.4 Effects of Intervention Fidelity

Focusing on SCIP classes only, the ICCs indicated significant between-class variation in the outcome variables at t2, even after controlling for t1 between-class differences ($ICC_{SA} = 0.105$, $p < 0.001$; $ICC_{SB} = 0.077$, $p < 0.01$; see Table S3). The results (Table 3) showed that the quantity component of intervention fidelity—i.e., the number of SCIP interventions—did not affect the outcome variables ($b_{SB} = 0.107$, $p > 0.05$; $b_{SA} = 0.077$, $p > 0.05$). In contrast, concept adherence—the quality component of intervention fidelity—had significant positive effects on both DPAS subtests ($b_{SB} = 0.216$, $p < 0.001$; $b_{SA} = 0.287$, $p < 0.001$). Notably, these effects accounted for a large share of between-class variance ($R^2_{SB} = 0.708$, $R^2_{SA} = 0.903$). Thus, between-class differences at t2 in the outcomes were largely due to the degree to which teachers implemented SCIP as intended.

5 Discussion

The current intervention study investigated whether the SCIP approach positively affected students' scientific creativity, specifically divergent problem-solving in science. Divergent problem-solving ability was assessed using two subscales of the DPAS [53]: SA and SB.

Overall, the results of this study support the idea that SCIP had the intended positive effects on students' divergent problem-solving in science. The effects were significant for both subtests (bs = 0.22 and 0.31) and accounted for 29% (SB) and 40% (SA) of the variance between classes. Excluding the two control classes that deviated from the CTRL conditions (i.e., increased program differentiation) resulted in higher R^2-values of about 48% for SB and SA. Moreover, the results also indicated significant variation in the outcomes between the SCIP classes. This variance could largely be attributed (71% SB, 90% SA) to the degree of concept adherence between teachers. Notably, the quantity component of intervention fidelity (i.e., the number of interventions) did not affect student outcomes. Taken together, this paper provides initial evidence regarding the effectiveness of the SCIP approach. Although, according to Cohens benchmarks [67], the effects would be considered small, recent work on the interpretation of effect sizes of educational interventions [68, 69] suggests that the effects reported in the current study are to be considered large. The benchmarks proposed by Kraft [69]—empirically derived from randomized controlled trails—consider effects of 0.20 (SD) or greater as large. As Kraft argues, the significance of a 0.2 SD effect of an intervention becomes highlighted when considering that one year of schooling is associated with a 0.4 SD increase in achievement measures and schools account for only about 40% of achievement gains. However, it must be considered that the proposed benchmarks refer to education interventions targeting math and reading skills. They are based on randomized controlled trials, which are known to result in smaller effect sizes than quasi-experimental studies [68] such as the current one. Therefore, future research should address the question what effect sizes are to be expected for interventions targeting scientific creativity.

Table 3 Effects of implementation fidelity on DPAS subtests

	SB		SA	
	M1	M2	M1	M2
	b (SE) [95%-CI]	b (SE) [95%-CI]	b (SE) [95%-CI]	b (SE) [95%-CI]
Within Class				
Pretest	0.612*** (0.064) [0.486, 0.738]	0.611*** (0.060) [0.493, 0.729]	0.556*** (0.052) [0.455, 0.658]	0.575*** (0.052) [0.473, 0.678]
GPA_MS	– 0.130* (0.053) [– 0.234, –0.025]	– 0.133** (0.051) [– 0.234, –0.033]	– 0.190*** (0.048) [– 0.284, –0.095]	– 0.191*** (0.047) [– 0.284, –0.098]
Gender (Male)	– 0.198* (0.092) [– 0.378, –0.017]	– 0.200* (0.090) [– 0.376, –0.024]	– 0.342*** (0.089) [– 0.516, –0.167]	– 0.326*** (0.088) [– 0.499, –0.153]
Grade	0.193 (0.167) [– 0.134, 0.520]	0.082 (0.127) [– 0.166, 0.331]	0.258 (0.164) [– 0.081, 0.597]	0.098 (0.101) [– 0.099, 0.296]
Residual Variance	0.421*** (0.053) [0.318, 0.525]	0.421*** (0.053) [0.318, 0.525]	0.405*** (0.043) [0.320, 0.490]	0.406*** (0.043) [0.320, 0.491]
Between Class				
Number of Intervention	0.107 (0.064) [– 0.018, 0.233]		0.077 (0.077) [– 0.074, 0.229]	
Concept adherence		0.216*** (0.055) [0.109, 0.323]		0.287*** (0.037) [0.214, 0.361]
Intercept	– 0.005 (0.052) [– 0.107, 0.096]	– 0.005 (0.039) [– 0.080, 0.071]	0.002 (0.061) [– 0.119, 0.122]	– 0.002 (0.032) [– 0.065, 0.062]
Residual Variance	0.053* (0.026) [0.003, 0.104]	0.019 (0.013) [– 0.006, 0.044]	0.084** (0.029) [0.026, 0.141]	0.009 (0.011) [– 0.012, 0.030]
R^2 Within	0.540	0.531	0.542	0.545
R^2 Between	0.206	0.708	0.082	0.903

CI = confidence interval. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

On average, teachers implemented 9.38 SCIP interventions throughout the school year. However, there was a large variation, ranging from four to 18 SCIP interventions. Regarding program differentiation, the question arised whether control classes could be regarded as CTRL. The CTRL condition was supported for 10 out of the 12 control classes. In two classes, teachers implemented some SCIP elements and thus somewhat deviated from the CTRL.

Providing a teacher-training program may have contributed to the effectiveness of SCIP [70]. To enhance the quality of the investigation, intervention fidelity was also assessed, focusing on the structural level, specifically the dosage and degree of adherence, as adherence to the specified intervention according to the program design is a major element of fidelity [55]. Intervention fidelity has gained increasing significance for research on evaluation or treatment effectiveness because it provides researchers with an understanding of the relationships between interventions and outcomes [71]. Greater fidelity increases internal and external validity, statistical power and effect size [50, 72–74]. In the intervention classes, the mean degree of adherence reached a high value of 91% (SD = 7); typically, 80–100% integrity is considered high [52, 72, 75]. It is evident that adequate training and supervision of interventionists are required to ensure a high degree of intervention fidelity [51, 52], and the high value of adherence in the present study can be explained by the fact that all teachers were involved in a teacher-training program on SCIP.

Nonetheless, the current study had some design features (e.g., non-randomized group assignment, use of researcher developed measures) that are known to be associated with higher effect sizes [68]. Thus, future research on the SCIP approach should employ randomization and alternative outcome measures (also covering other aspects of scientific creativity). Moreover, although this study assessed dosage and degree of adherence, other aspects of intervention fidelity were not assessed and should be considered in future research. The somewhat unexpected insignificance of intervention dosage should also be considered in future SCIP research.

6 Conclusion and outlook

Scientific creativity will play an increasingly important role in science education, especially as this way of thinking is perceived as a prerequisite for both academic and professional success [76]. In the present study, the SCIP program enhanced divergent problem-solving skills in adolescents and, thus, may play a significant role in enhancing scientific creativity.

In addition to the tools used in this study, the SCIP program includes other interventions and reflective tools that promote the most important skills needed for effective scientific creativity. These include divergent thinking and problem-solving skills, as well as bisociation imagination, analogy and metacognition. Thus, future work should look at how each tool in the SCIP program affects the various aspects of scientific creativity. It can be assumed that metacognition plays a particularly large role in promoting scientific creativity, which should be investigated in future studies.

The SCIP program has shown to be easy to integrate into school lessons, as teachers only need to make minor changes to their teaching. This was supported by the non-significant effect of dosage on effectiveness in the present study. Thus, it appears that only a few well-implemented interventions are needed to achieve significant increases in divergent problem-solving skills.

Note: Selected worksheets on the SCIP techniques can be downloaded from the homepage www.school-creative-solutions.at or requested from the authors.

Author Contributions K.H.: Kurt Haim/W.A.: Wolfgang Aschauer / C.W.: Christoph Weber All authors contributed to the study conception and design. Material preparation and data collection were performed by K.H. and W.A. / Data analysis was performed by C.W. / The introduction and theoretical framework of the first draft were written by K.H. / Methodology and Results of the first draft were written by W.A. and C.W. / All authors read and approved the final manuscript.

Funding No funds, grants, or other support was received.

Availability of data and materials The datasets and materials are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the institutional review board of the Upper Austrian Department of Education. Consent to participate was obtained from all individual participants included in the study.

Consent for publication Not applicable, because we obtained only anonymized data.

Informed consent Informed consent was obtained from all individual participants included in the study.

Competing Interests The authors declare no competing interests.

References

1. Rieckmann M, Gardiner S, Mindt L. Education for Sustainable Development Goals: learning objectives. Paris: UNSESCO; 2017. https://unesdoc.unesco.org/ark:/48223/pf0000247444
2. Kind PM, Kind V. Creativity in science education: perspectives and challenges for developing school science. Stud Sci Educ. 2007;43:1–37. https://doi.org/10.1080/03057260708560225.
3. Marope M, Griffin P, Gallagher C. Future competences and the future of curriculum. 2017. http://www.ibe.unesco.org/en/news/document-future-competences-and-future-curriculum
4. Newton LD, Newton DP. Creativity in 21st-century education. Prospects. 2014;44:575–89. https://doi.org/10.1007/s11125-014-9322-1.

5. Ayas MB, Sak U. Objective measure of scientific creativity: psychometric validity of the creative scientific ability test. Thinking Skills Creativity. 2014;13:195–205. https://doi.org/10.1016/j.tsc.2014.06.001.
6. Hadzigeorgiou Y, Fokialis P, Kabouropoulou M. Thinking about creativity in science education. CE. 2012;03:603–11. https://doi.org/10.4236/ce.2012.35089.
7. Hu W, Adey P. A scientific creativity test for secondary school students. Int J Sci Educ. 2002;24:389–403. https://doi.org/10.1080/09500690110098912.
8. Hu W, Shi QZ, Han Q, Wang X, Adey P. Creative scientific problem finding and its developmental trend. Creat Res J. 2010;22:46–52. https://doi.org/10.1080/10400410903579551.
9. Pacheco SC, Herrera IC. A conceptual proposal and operational definitions of the cognitive processes of complex thinking. Thinking Skills Creativity. 2021;39: 100794. https://doi.org/10.1016/j.tsc.2021.100794.
10. Sternberg RJ, Todhunter RJE, Litvak A, Sternberg K. The relation of scientific creativity and evaluation of scientific impact to scientific reasoning and general intelligence. J Intell. 2020. https://doi.org/10.3390/jintelligence8020017.
11. Feist GJ. The Function of Personality in Creativity: The Nature and Nurture of the Creative Personality. In: Kaufman JC, Sternberg RJ, editors. The Cambridge Handbook of Creativity. New York: Cambridge university press; 2010. p. 113–130. https://doi.org/10.1017/CBO9780511763205
12. Hu W, Wu B, Jia X, Yi X, Duan C, Meyer W, Kaufman JC. Increasing students' scientific creativity: the "learn to think" intervention program. J Creat Behav. 2013;47:3–21. https://doi.org/10.1002/jocb.20.
13. Huang P-S, Peng S-L, Chen H-C, Tseng L-C, Hsu L-C. The relative influences of domain knowledge and domain-general divergent thinking on scientific creativity and mathematical creativity. Thinking Skills Creativity. 2017;25:1–9. https://doi.org/10.1016/j.tsc.2017.06.001.
14. OECD. PISA 2025 Science framework. 2023. https://pisa-framework.oecd.org/science-2025/assets/docs/PISA_2025_Science_Framework.pdf.
15. Yang K-K, Lee L, Hong Z-R, Lin H-s. Investigation of effective strategies for developing creative science thinking. International Journal of Science Education. 2016;38:2133–51. https://doi.org/10.1080/09500693.2016.1230685
16. Haim K, Aschauer W. Fostering Scientific Creativity in the Classroom: The Concept of Flex-Based Learning. International Journal of Learning, Teaching and Educational Research. 2022;21:196–230. https://doi.org/10.26803/ijlter.21.3.11
17. Runco MA, Jaeger GJ. The standard definition of creativity. Creat Res J. 2012;24:92–6. https://doi.org/10.1080/10400419.2012.650092.
18. Simonton DK. Scientific creativity as constrained stochastic behavior: the integration of product, person, and process perspectives. Psychol Bull. 2003;129:475–94. https://doi.org/10.1037/0033-2909.129.4.475.
19. Feist GJ. Creativity in science. In: Runco MA, Pritzker SR, editors. Encyclopedia of creativity. 2nd ed.: Academic Press/Elsevier; 2011. p. 296–302.
20. Beghetto RA. Ideational code-switching: walking the talk about supporting student creativity in the classroom. Roeper Rev. 2007;29:265–70. https://doi.org/10.1080/02783190709554421.
21. Beghetto RA, Kaufman JC. Toward a broader conception of creativity: a case for "mini-c" creativity. Psychol Aesthet Creat Arts. 2007;1:73–9. https://doi.org/10.1080/02783190709554421.
22. Boden MA. The creative mind: myths and mechanisms. New York, London: Routledge; 2004.
23. Bonetto E, Pichot N, Pavani J-B, Adam-Troïan J. Creative individuals are social risk-takers: Relationships between creativity, social risk-taking and fear of negative evaluations. Creativity. Theor Rese-Appl. 2021;7:309–20. https://doi.org/10.2478/ctra-2020-0016.
24. Runco MA, Acar S. Divergent thinking as an indicator of creative potential. Creat Res J. 2012;24:66–75. https://doi.org/10.1080/10400419.2012.652929.
25. Runco MA. Divergent thinking. In: Runco MA, Pritzker SR, editors. Encyclopedia of creativity, vol. 1. San Diego, California: Academic Press; 1999. p. 577–82.
26. Kaufman JC, Plucker JA, Baer J. Essentials of creativity assessment. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2008.
27. Ward TB, Smith SM, Vaid J. Conceptual structures and processes in creative thought. In: Ward TB, Smith SM, Vaid J, editors. Creative Thought: An investigation of conceptual structures and processes. Washington, DC: American Psychological Association; 1997. p. 1–27. https://doi.org/10.1037/10227-000
28. Benedek M, Jurisch J, Koschutnig K, Fink A, Beaty RE. Elements of creative thought: Investigating the cognitive and neural correlates of association and bi-association processes. Neuroimage. 2020;210: 116586. https://doi.org/10.1016/j.neuroimage.2020.116586.
29. Mednick S. The associative basis of the creative process. Psychol Rev. 1962;69:220. https://doi.org/10.1037/h0048850.
30. Condell J, Wade J, Galway L, McBride M, Gormley P, Brennan J, Somasundram T. Problem solving techniques in cognitive science. Artif Intell Rev. 2010;34:221–34. https://doi.org/10.1007/s10462-010-9171-0.
31. Taylor M. Imagination. In: Runco MA, Pritzker SR, editors. Encyclopedia of creativity. 2nd ed.: Academic Press/Elsevier; 2011. 637–643.
32. Magno C. Explaining the creative mind. Int J Res Rev. 2009;3:10–9.
33. Flavell JH. Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. Am Psychol. 1979;34:906. https://doi.org/10.1037/0003-066X.34.10.906.
34. Schraw G, Crippen KJ, Hartley K. Promoting self-regulation in science education: metacognition as part of a broader perspective on learning. Res Sci Educ. 2006;36:111–39. https://doi.org/10.1007/s11165-005-3917-8.
35. Schraw G. Promoting general metacognitive awareness. Instr Sci. 1998;26:113–25. https://doi.org/10.1023/A:1003044231033.
36. Kaufman JC, Beghetto RA. In praise of clark kent: creative metacognition and the importance of teaching kids when (Not) to be creative. Roeper Rev. 2013;35:155–65. https://doi.org/10.1080/02783193.2013.799413.
37. Lin C, Hu W, Adey P, Shen J. The influence of CASE on scientific creativity. Res Sci Educ. 2003;33:143–62. https://doi.org/10.1023/A:1025078600616.
38. Willerson A, Mullet DR. Creativity research in education from 2005–2015: A systematic review and synthesis. The International Journal of Creativity and Problem Solving. 2017;27:5–22. https://link.gale.com/apps/doc/A533556208/AONE?u=anon~2241f4c9&sid=googleScholar&xid=f55cf281
39. Bi H, Mi S, Lu S, Hu X. Meta-analysis of interventions and their effectiveness in students' scientific creativity. Thinking Skills Creativity. 2020;38: 100750. https://doi.org/10.1016/j.tsc.2020.100750.
40. Aktamis H, Ergin Ö. The effect of scientific process skills education on students'scientific creativity, science attitudes and academic achievements. Asia-Pacific Forum on Science Learning and Teaching. 2008;9. https://www.eduhk.hk/apfslt/

41. Siew NM, Chin MK. Design, Development and evaluation of a problembased with cooperative module on scientific creativity of preschoolers. Journal of Baltic Science Education. 2018;17:215–28. http://www.scientiasocialis.lt/jbse/files/pdf/vol17/215-228.Siew_JBSE_Vol.17_No.2.pdf
42. Turnbull M, Littlejohn A, Allan M. Creativity and collaborative learning and teaching strategies in the design disciplines. Ind High Educ. 2010;24:127–33. https://doi.org/10.5367/000000010791191029.
43. Nasiri Shm. Computer application in teaching physics and its effect on general and scientific creativity in relation to intelligence, SES, and attitude towards media: Panjab University; 2000. https://osre.ncert.gov.in/abstract/Nasiri_S_H_M/551
44. Lawson AE. The nature and development of scientific reasoning: a synthetic view. Int J Sci Math Educ. 2004;2:307–38. https://doi.org/10.1007/s10763-004-3224-2.
45. Sun M, Wang M, Wegerif R. Effects of divergent thinking training on students' scientific creativity: the impact of individual creative potential and domain knowledge. Thinking Skills and Creativity. 2020. https://doi.org/10.1016/j.tsc.2020.100682.
46. Sidek R, Halim L, Buang NA, Arsad NM. Fostering scientific creativity in teaching and learning science in schools: A systematic review. Jurnal Penelitian dan Pembelajaran IPA. 2020;6:13–35. https://doi.org/10.30870/jppi.v6i1.7149
47. Syukri M, Halim L, Mohtar LE, Soewarno S. The impact of engineering design process in teaching and learning to enhance studentsâ€™ science problem-solving skills. Jurnal Pendidikan IPA Indonesia. 2018; 7: 66–75. https://doi.org/10.15294/jpii.v7i1.12297
48. Lyman FT. The responsive classroom discussion: the inclusion of all students. In: Anderson AS, editor. Mainstreaming digest: a collection of faculty and student papers. Maryland: University of Maryland; 1981. p. 109–13.
49. Haim E, Stella M. Cognitive networks for knowledge modelling: A gentle tutorial for data-and cognitive scientists. 2023. https://doi.org/10.31234/osf.io/dsmuk.
50. Busch C, Dreyer R, Janneck M. Blended health coaching for work-linked couples: coaches' intervention fidelity and empathy matter! coaching. Theor Prax. 2022. https://doi.org/10.1365/s40896-022-00065-9.
51. Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: are implementation effects out of control? Clin Psychol Rev. 1998;18:23–45. https://doi.org/10.1016/S0272-7358(97)00043-3.
52. Gearing RE, El-Bassel N, Ghesquiere A, Baldwin S, Gillies J, Ngeow E. Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. Clin Psychol Rev. 2011;31:79–88. https://doi.org/10.1016/j.cpr.2010.09.007.
53. Aschauer W, Haim K, Weber C. A contribution to scientific creativity: a validation study measuring divergent problem solving ability. Creat Res J. 2021;34:195–212. https://doi.org/10.1080/10400419.2021.1968656.
54. Snyder A, Mitchell J, Bossomaier T, Pallier G. The creativity quotient: an objective scoring of ideational fluency. Creat Res J. 2004;16:415–9. https://doi.org/10.1080/10400410409534552.
55. Mowbray CT, Holter MC, Teague GB, Bybee D. Fidelity criteria: development, measurement, and validation. Am J Eval. 2003;24:315–40. https://doi.org/10.1177/109821400302400303.
56. Case LP, Speece DL, Silverman R, Ritchey KD, Schatschneider C, Cooper DH, et al. Validation of a supplemental reading intervention for first-grade children. J Learn Disabil. 2010;43:402–17. https://doi.org/10.1177/0022219409355475.
57. Kamps D, Abbott M, Greenwood C, Wills H, Veerkamp M, Kaufman J. Effects of small-group reading instruction and curriculum differences for students most at risk in kindergartenL: two-year results for secondary—and tertiary-level interventions. J Learn Disabil. 2008;41:101–14. https://doi.org/10.1177/0022219407313412.
58. O'Donnell CL. Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. Rev Educ Res. 2008;78:33–84. https://doi.org/10.3102/0034654307313793.
59. Nelson MC, Cordray DS, Hulleman CS, Darrow CL, Sommer EC. A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. J Behav Health Serv Res. 2012;39:374–96. https://doi.org/10.1007/s11414-012-9295-x.
60. Neuschmidt O, Barth J, Hastedt D. Trends in gender differences in mathematics and science (TIMSS 1995–2003). Stud Educ Eval. 2008;34:56–72. https://doi.org/10.1016/j.stueduc.2008.04.002.
61. Muthén LK, Muthén BO. Mplus user's guide. Eight. Los Angeles: Muthén & Muthén; 1998.
62. Bollen KA. Structural equations with latent variables. 210th ed. New York: John Wiley & Sons; 1989.
63. Lüdtke O, Marsh HW, Robitzsch A, Trautwein U. A 2 × 2 taxonomy of multilevel latent contextual models: accuracy-bias trade-offs in full and partial error correction models. Psychol Methods. 2011;16:444–67. https://doi.org/10.1037/a0024376.
64. Morin AJS, Marsh HW, Nagengast B, Francesca Scalas L. Doubly latent multilevel analyses of classroom climate: an illustration. J Exp Educ. 2014;82(2):143–67. https://doi.org/10.1080/00220973.2013.769412.
65. Bhaktha N, Lechner CM. To score or not to score? A simulation study on the performance of test scores, plausible values, and SEM, in regression with socio-emotional skill or personality scales as predictors. Front Psychol. 2021;12: 679481.
66. Asparouhov T, Muthén BO. Plausible values for latent variables using Mplus. Los Angeles, CA: Mplus Technical Appendix; 2010.
67. Cohen J. A power primer. Psychol Bull. 1992;112:155–9.
68. Cheung ACK, Slavin RE. How methodological features affect effect sizes in education. Educ Res. 2016;45:283–92. https://doi.org/10.3102/0013189X16656615.
69. Kraft MA. Interpreting effect sizes of education interventions. Educ Res. 2020;49:241–53. https://doi.org/10.3102/0013189X20912798.
70. Cheung A, Slavin RE, Kim E, Lake C. Effective secondary science programs: a best-evidence synthesis. J Res Sci Teach. 2017;54:58–81. https://doi.org/10.1002/tea.21338.
71. Borrelli B, Sepinwall D, Ernst D, Bellg AJ, Czajkowski S, Breger R, et al. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. J Consult Clin Psychol. 2005;73:852–60. https://doi.org/10.1037/0022-006X.73.5.852.
72. Biel CH, Buzhardt J, Brown JA, Romano MK, Lorio CM, Windsor KS, et al. Language interventions taught to caregivers in homes and classrooms: a review of intervention and implementation fidelity. Early Childhood Res Quarte. 2020;50:140–56. https://doi.org/10.1016/j.ecresq.2018.12.002.
73. Dusenbury L, Brannigan R, Falco M, Hansen WB. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. Health Educ Res. 2003;18:237–56. https://doi.org/10.1093/her/18.2.237.
74. Lee Y-J, Chue S. The value of fidelity of implementation criteria to evaluate school-based science curriculum innovations. Int J Sci Educ. 2013;35:2508–37. https://doi.org/10.1080/09500693.2011.609189.

75. Perepletchikova F, Kazdin AE. Treatment integrity and therapeutic change: issues and research recommendations. Clin Psychol Sci Pract. 2005;12:365–83. https://doi.org/10.1093/clipsy/bpi045.
76. Gajda A, Karwowski M, Beghetto RA. Creativity and academic achievement: a meta-analysis. J Educ Psychol. 2017;109:269–99. https://doi.org/10.1037/edu0000133.