Integrated STEM Education: The Effects of a Long-Term Intervention on Students' Cognitive Performance

Haydée De Loof [1,2*], Jelle Boeve-de Pauw [1,3], Peter Van Petegem [3]

[1] *University of Antwerp*, BELGIUM
[2] *KdG University of Applied Sciences and Arts*, BELGIUM
[3] *Utrecht University*, THE NETHERLANDS

***Corresponding Author:** Haydee.DeLoof@kdg.be

ABSTRACT

Integrated Science, Technology, Engineering and Mathematics (iSTEM) education is a promising approach to attracting more qualified and better motivated students to STEM fields. In this study, we respond to one of the salient challenges facing integrated STEM educational research, namely investigating its educational impact. We developed a large-scale intervention where all STEM components are integrated and examined the impact this integrated STEM curriculum has had on cognitive performances regarding physics, mathematics, technological concepts, and integrated physics and mathematics. In total, 859 grade 9 students, distributed across 39 Flemish different schools, participated in a long-term study. The results of multilevel analyses show that iSTEM education had positive effects on cognitive performance in terms of mathematics knowledge and application and technological concepts. Also, differential intervention effects were found with regard to student characteristics. Since the impact was only apparent after two years, we stress the importance of a long-term integrated STEM approach.

Keywords: STEM, integrated STEM, cognitive performance, longitudinal research

INTRODUCTION

Generating a sufficient number of qualified professionals in science, technology, engineering and mathematic (STEM) areas is a matter of international concern (Thibaut et al., 2018a; Hernandez et al., 2014; Bøe, Henriksen, Lyons and Schreiner, 2011). Awareness of the problem with regard to young people's increasing reluctance to participate in STEM emerged in the early 1990s, and this has been a growing problem to this day (Bøe et al., 2011; Moore and Smith, 2014; Keith, 2018; Hermans et al., 2022; De Loof et al., 2022), as national reports continue to identify shortages of STEM graduates. The World Economic Forum (2016) predicted an increased demand for specialists in the STEM field for the years to come and stated that the pace of technology adoption is expected to remain unabated and may accelerate in some areas (World Economic Forum, 2020). The current knowledge-based society demands a large number of students graduating from STEM-related fields (National Academies of Science, 2007). Countries need a sound economy and need to find solutions for societal and environmental matters, such as sustainable energy production in a world with shrinking resources, adequate healthcare in an aging society, and well-considered technology development (Wang, Moore, Roehrig and Park, 2011; Kjærnsli and Lie, 2011; Bøe et al., 2011). Integrated STEM can play a central role in motivating students to choose a STEM study or profession

and has the potential to improve students' learning (Honey, Pearson and Schweingruber, 2014; Becker and Park, 2011).

Integrated STEM

STEM is an integration of the four subjects: science, technology, engineering, and mathematics (Wang et al., 2011). However, as the term is widely used, there is no consensus about the nature and range of the concept. Some researchers and educators use the term STEM to refer to one or more of its components, others use it only in the integrated sense (Wang et al., 2011; English, 2016). As we are discussing the term in the integrated sense, we will use 'integrated STEM' (iSTEM) in this paper. Sanders (2009) defines iSTEM approaches as "Approaches that explore teaching and learning between/among any two or more of the STEM subject areas, and/or between a STEM subject and one or more other school subjects" (Sanders, 2009: 21). According to Honey et al. (2014), iSTEM education includes a range of different experiences that involve some degree of connection. "The experiences may occur in one or several class periods, throughout a curriculum, be reflected in the organization of a single course or an entire school or be encompassed in an out-of-school activity" (Honey et al., 2014: 2). Consequently, they define integration as "…working in the context of complex phenomena or situations on tasks that require students to use knowledge and skills from multiple disciplines" (Honey et al., 2014: 52).

In the current study, we use Honey et al.'s definition, and thus we approach iSTEM in terms of the integration of all its components into a single curricular project that emphasizes concepts and their application from across the four disciplines (Roehrig, Moore, Wang and Park, 2012). Within this approach, the literature differentiates between multidisciplinary and interdisciplinary integration (Wang et al., 2011). The metaphor of chicken noodle soup versus tomato soup provided by Lederman and Niess (1997), is often used to explain the differences between these two forms of integration. The chicken noodle soup represents multidisciplinary integration, where each ingredient maintains its identity without a direct mixture in the totality of the integration. Multidisciplinarity starts from subject-based content and skills, and students are expected to form connections between the subjects that they have been taught in different classes (Wang et al., 2011). Tomato soup, on the other hand, represents interdisciplinary integration, where the boundaries between subjects are blurry. Interdisciplinarity starts from a problem that requires an understanding of the content and skills of multiple subjects (Wang et al., 2011). Vasquez, Sneider, and Comer (2013) add an additional level of increased integration by introducing the concept of transdisciplinary integration. Knowledge and skills from multiple disciplines are hereby applied to solve real-world problems. In the current study, we approach education in iSTEM as a transdisciplinary concept.

Educational Research in Integrated STEM

Removing barriers between disciplines is meant to increase students' conceptual understanding and achievements regarding STEM topics and increase recognition of the relevance of the subjects in relation to each other and to the context of real-world problems (Honey et al., 2014; Thibaut et al., 2018a). Integrated STEM education is a promising approach to attracting more qualified and motivated students in STEM fields by improving students' interest and learning in STEM. It has received increasing attention from educators and researchers over the past decade (Honey et al., 2014; Yang and Baldwin, 2020). Besides the possible positive effects of iSTEM education on the general student population, it has also been argued that iSTEM might be particularly beneficial to certain student populations. Cantrell, Pekca, and Ahmad (2006) for instance, showed that an integrated engineering curriculum diminished achievement gaps in typically low-achieving ethnic minority student groups. Newton et al. (2020) demonstrated that informal STEM learning in robotics and game design could influence computational thinking skills in African-American students living in an urban context. Other studies demonstrate that gender differences in performance might reduce when students follow iSTEM courses linked with real-world activities (Standish, Christensen, Knezek, Kjellstorm and Bredder, 2016). Hence, student characteristics that might have an effect on cognitive STEM outcomes might have a differential impact in an iSTEM educational approach. In the literature, such characteristics are well documented: previous research has indicated that gender, abstract reasoning ability and socioeconomic status (SES) might influence cognitive scores on STEM domains (e.g., Halpern et al., 2007; Deary, Strand, Smith and Fernandes, 2007; Yerdelen-Damar and Peşman, 2013).

To conclude, an iSTEM educational approach is promising for both the general student population and for a variety of students with different characteristics. In response, numerous new teaching materials, projects, and even complete study programs have been developed (e.g., Yang et al., 2021). Such a development entails the challenge to investigate empirical evidence to support the effective implementation of iSTEM education (Becker and Park, 2011). Indeed, the notion that learning becomes more meaningful and prolonged when students can make connections between STEM concepts has prompted research that aims to investigate the cognitive benefits of iSTEM education.

Becker and Park (2011) have synthesized research findings on the effects of integrated approaches among STEM subjects on students' cognitive performances. In their meta-analysis they described 28 studies reporting on effectiveness regarding students' learning in integrated STEM conditions. According to Becker and Park (2011), the small number of studies is due to the finding that many pieces of research are in the form of opinion papers without empirical data. Studies varied in the degree to which they addressed the integration of two or more STEM-subjects, the number of participants, and their age. A first gap in the current body of knowledge is the number of studies that integrated all components of STEM and reported on all associated cognitive outcomes. Only one study addressed the integration of all components, i.e., a study on the effect of integrated STEM on students with learning disabilities (Lam et al., 2008). Five studies discussed achievement scores after integration of S-T-E, and five studies reported on scores after the integration of S-T-M. Other studies integrated only two components. Regarding the measured achievement, only Lam et al. (2008), reported on the scores on all components, and just two studies measured the scores on S-T-E. No studies reported scores on questions addressing integrated STEM. A second concern is the low number of participants and the small scale of the interventions. Since the mean number of participants is 174.58 (min. = 21; max. = 1,053), it is difficult to draw far-reaching conclusions. A third shortage is that studies are limited in terms of time perspective. No longitudinal studies could be included, which has the implication that little is known about the long-term effect of iSTEM education.

Studies published after Becker and Park's (2011) meta-analysis encounter the same problems (i.e., a skewed focus on science at the expense of mathematics, no integration of all subjects, limited numbers of participants, and no studies from a long-term perspective) (English, 2016; Yildirim, 2016), and continue to be small in number. To conclude, long-term research with all STEM components integrated is very rare and, as a result, the effects of an iSTEM approach on cognitive performances is a crucial gap in the field. More empirical research on the educational effects of (integrated) STEM education is therefore needed (Honey et al., 2014). With our current long-term study, we respond to this challenge and to the need to fill the gaps in integrated STEM educational research. We focus explicitly on the effect of a large-scale intervention where all STEM components are integrated in the developed learning modules.

Design of the Intervention

The intervention, called STEM@School is a collaborative project between two universities (KU Leuven, University of Antwerp) and two educational umbrella organizations (GO!, Catholic Education Flanders) covering approximately 70% of all schools in Flanders. The KU Leuven developed the learning materials in collaboration with teacher design teams, and the University of Antwerp evaluated the project. The role of the two umbrella organizations was to support the participating schools in their implementation, and to monitor the content of the developed materials so that they cover all learning objectives and curriculum guidelines.

Five iSTEM learning modules were developed. Schools incorporated three of these modules into the curriculum in grade 9 (third year in secondary education; age: 13-15 years), and two in grade 10 (fourth year in secondary education; age: 14-16 years). The participating schools introduced an integrated STEM subject in which the learning modules were addressed. To implement these learning modules, 4 to 5 teaching hours a week were required for the duration of each of two semesters. The schools taught the integrated STEM subject partly within the teaching hours of the regular mathematics, physics and engineering classes (where the regular content was aligned with the curriculum of the integrated STEM subject), and partly within additional hours in the form of optional classes. More detailed information of the project and its implementation approach can be found in the project paper of STEM@School (Knipprath et al., 2018).

The learning modules consisted of challenges that were relevant in terms of societal and ecological problems, applying a transdisciplinary approach (Vasquez et al., 2013). Students address these challenges by applying knowledge and skills across disciplines, hereby making connections between principles and concepts. Problem-solving in an integrated STEM context also requires inquiry and design competences on the part of the students (Thibaut et al., 2018a). These characteristics constituted the core of the iSTEM intervention and were the foundation of all learning modules.

An example of one of the learning modules is the challenge of the optimization of traffic flow through a green wave (i.e., the coordination of traffic lights to allow continuous traffic flow). Students had to design and program a car in such a way that it could drive through a green wave without exceeding a safe speed limit. To succeed in this challenge, they had to use knowledge and skills from all STEM disciplines, such as acceleration (science), building the car with appropriate materials (technology), programming the car (engineering), and functions (mathematics). Obviously, this division is to some extent artificial, as these domains are interdependent. For instance, mathematics is already embedded in the physical concept of acceleration (Becker and Park, 2011), and some authors consider engineering as a subset of technology (Williams, 2011). Nevertheless, all modules could be considered as challenges which incorporated themes from the different STEM domains.

Current Study

Given the need for long-term educational research regarding iSTEM education, we aimed to evaluate the effectiveness of a large-scale two-year intervention in which students had to respond to relevant challenges by making use of knowledge and skills from different STEM domains. To respond to challenges posed in previous integrated STEM educational research, we incorporated all four domains in the intervention, and investigated the cognitive effects on physics' knowledge, physics' application, mathematics' knowledge, mathematics' application, technological concepts, and integrated physics and mathematics. We put forward two research questions:

Research question 1. What is the impact of an iSTEM curriculum on cognitive performances regarding physics (both knowledge and application), mathematics (both knowledge and application), technological concepts, and integrated physics and mathematics after one and two years?

Research question 2. What is the differential effectiveness of the iSTEM curriculum regarding student characteristics (i.e., gender, SES and abstract reasoning)?

METHOD

Participants and Procedure

The schools in this study were part of STEM@School and volunteered to take part in this longitudinal study. Thirty schools involving 612 grade 9 students implemented the experimental condition of the iSTEM education program. To assemble a representative control group, all Flemish schools (i.e., schools serving the Dutch-speaking community of Belgium) were listed, and an inventory of relevant characteristics was created, such as the number of students, study track options, and membership of educational umbrella organizations. Subsequently, for each experimental school, three matching schools were selected at random and invited to participate in the project. Control schools were invited through a letter, and if no response was received, school administrators were called by a researcher as a follow-up. Nine control schools took part in the project, involving 247 students in the control condition of a traditional education program, with separate physics, engineering, and mathematics courses.

The students in this study were taking classes in one of the following three study tracks: 1. Science and Mathematics, 2. Engineering, and 3. Latin and Mathematics. The total number of participants and the division over condition and study track can be found in **Table 1**.

Table 1. Number of participants (absolute and relative) divided over condition and study track

	Experimental condition	Control condition	Total	%
Science & mathematics	396	169	565	66%
Engineering	201	47	248	29%
Latin & mathematics	15	31	46	5%
Total	**612**	**247**	**859**	100%
%	71%	29%	100%	100%

The participants totalled 859 (66% boys and 34% girls) grade 9 students with a mean age of 13.86 years ($SD =$ .54) at the start of the study. We followed a quasi-experimental longitudinal design over two years. Three measurement moments were undertaken in both the experimental and control conditions: (1) before the start of grade 9, (2) at the end of grade 9, and (3) at the end of grade 10 (**Figure 1**).
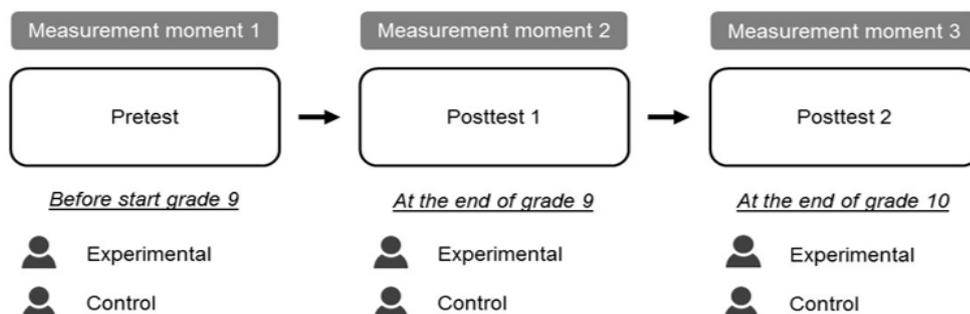

Figure 1. Measurement moments before start of grade 9, after the end of grade 9, and after the end of grade 10

While in total 859 unique participants were involved in the study, some of them were missing at different measurement moments. This could be due to schools dropping out of the project over time, or the failure of

schools to administer surveys to students on a measurement moment, and because of the illness of individual students on a particular measurement moment. No selective attrition was observed, as Little's MCAR test showed that the data were missing completely at random. **Table 2** provides an overview of the number of recorded responses of students over the three measurement moments.

Table 2. Number of recorded responses over measurement moments

	Time 1	Time 2	Time 3
Experimental	556	443	302
Control	238	208	138
Total	**794**	**651**	**440**

Students were allotted a unique code to guarantee their anonymity and to allow the researchers both to connect different questionnaires and tests within measurement moments, and to link questionnaires and tests across time. At the first measurement moment, students filled in an online questionnaire to provide demographic information and completed a test measuring abstract reasoning ability. Online multiple-choice tests were administered, measuring cognitive outcomes with regard to STEM concepts. Cognitive outcomes were re-assessed at the second and third measurement moments, with tests that were adapted to the expected level at the end of grades 9 and 10 respectively. Students completed the online questionnaires and tests during normal school hours under supervision of the schools' contact person of [name project]. Students and their parents were provided with information about the aim of the study, and with a passive informed consent procedure. This procedure was approved by an institutional ethical committee.

Instruments

Demographic information

Information regarding age, gender, and the SES of participants was obtained from the self-report of students on an online questionnaire. SES was determined by language spoken at home, country of birth of respondents and both parents (Tate, 1997), both parents' education, and both parents' occupational status[1] (Bornstein and Bradley, 2003). We performed exploratory factor analysis with varimax rotation on the above-mentioned variables, which led to two underlying variables: (1) origin and (2) occupation and education. The weighted sum of the factor scores on these two variables led to a total SES score for each student.

Abstract reasoning ability

We gathered information on abstract reasoning ability as a proxy for general and non-verbal intelligence (Conway, Cowan, Bunting, Therriault and Minkoff, 2002; Raven, Court and Raven, 1977). The test consisted of 40 items and involved inductive reasoning about spatial features and relationships. Every item consisted of a series of figures with one inconsistent figure.

Cognitive outcomes regarding STEM concepts

Six instruments were developed to measure cognitive performance with regard to physics, mathematics and technological concepts: (1) physics knowledge, (2) physics application, (3) mathematics knowledge, (4) mathematics application, (5) technological concepts, and (6) integrated physics and mathematics (IPM). Adapted instruments for these outcomes were developed to respond to the expected level at each measurement moment, i.e., the start and end of the ninth grade, and the end of the tenth grade. Instruments were constructed based on the curriculum for physics, mathematics, and technological concepts of the ninth and tenth grades by pedagogical and subject-matter experts. In the case of integrated physics and mathematics, no items were developed for measurement moment 1, as students at the beginning of grade 9 were not yet familiar with curricular mathematics and physics concepts that lend themselves to be integrated in an overarching question. Information about the number of items per instrument, and an example item for each of the six measured outcomes, can be found in **Appendix A**. To reduce the burden on students and to make it possible to administer these tests during school hours, only eight items of each instrument were selected at random by the online software and presented to the students.

---

[1] Typically, information about the family's economic situation is also used to calculate a measure for SES (Bornstein and Bradley, 2003). However, as we questioned students with an age range of 13-15, the current study did not include economic questions (Bradley and Corwyn, 2002).

The psychometric qualities of the tests for physics knowledge, physics application, mathematics knowledge, mathematics application, technological concepts, and integrated physics and mathematics were investigated, using latent trait models under Item Response Theory (IRT; Rizopoulos, 2006). A detailed description of the IRT analyses can be found in **Appendix B**. After the psychometric qualities of the instrument were investigated, a factor score for each student was calculated. This procedure was repeated for every cognitive test instrument for each of the three measurement moments. Due to poor psychometric qualities of the physics knowledge test at measurement moment 3, no individual scores were available for that scale.

Plan of Analysis

First, we investigated the intercorrelations among the dependent variables of the study which are shown in **Table 3**. Given that the correlations were between .01 (no linear relationship) and .42 (a small linear relationship), we conducted separate univariate analyses for all six cognitive outcomes.

Table 3. Intercorrelations among the dependent variables of the study

Variables	1.	2.	3.	4.	5.	6.
1. Physics knowledge						
2. Physics application	.11***					
3. Mathematics knowledge	.11***	.24***				
4. Mathematics application	.21***	.20***	.42***			
5. Technological concepts	-.03	-.01	.06*	.11***		
6. Integrated physics and mathematics	.10***	.06	.09**	.18**	.09***	

Note. The scores on the variables are aggregated IRT scores over time.
*p<.05. **p<.01. ***p<.001.

Subsequently, we constructed mixed models which allowed us to investigate the general and differential effects of the iSTEM intervention. We conducted multilevel analysis employing JMP software (John's Macintosh Project) version JMP pro 13. Linear mixed models in JMP make use of all data (and not only complete cases), thereby also including information of cases with missing values.

This study used a three-level model where measurement moments at level 1 were nested within students at level 2, which in turn were nested within schools at level 3. Multilevel modelling allows data to be clustered in groups (e.g., multiple measurement moments of one student, multiple students in one class group, multiple class groups in one school, etc.), and to have a hierarchical structure (e.g., students are part of a class group, and class groups are in turn part of a school). As students are measured three times, and their results are not independent. 'Student ID' was thus included as a random factor. Also, students learn together in a school, which could cause the outcomes of students within the same school to be more highly correlated than the outcomes of students between schools. Therefore, school was also included as a random factor. For all six investigated outcomes, we inspected whether a model with a fixed slope (random intercept model) fitted better to the data than a model with a random slope (random intercept and random slope model) (Raudenbush and Bryk, 2002). A multivariate likelihood-ratio test (2log (likelihood random slope)–2 log (likelihood random intercept)) revealed that the random slope model fitted better than the restricted (i.e., fixed slope) model in the case of physics application and mathematics knowledge. To examine agreement among students and agreement among schools we computed intra-cluster correlation coefficients (ICC).

With regard to the fixed effects, we included six main effects to control for their direct influence on the cognitive outcomes. Besides condition (0 = control condition, 1 = experimental condition), and measurement moment (1 = time 1, 2 = time 2, 3 = time 3), we also controlled for gender (1 = male, 2 = female), abstract reasoning and SES, as previous research indicated that these variables might influence cognitive scores on STEM domains (Halpern et al., 2007; Deary, Strand, Smith and Fernandes, 2007; Yerdelen-Damar and Peşman, 2013). Scores for abstract reasoning abilities and SES were standardized. It was also important to control for study (1= focus on science and mathematics, 2 = focus on engineering, 3 = focus on Latin and mathematics) as this variable was not homogenous in our sample.

To investigate the general intervention effects over time (see research question 1), we added the interaction between condition and measurement moment in the model. Differential intervention effects (see research question 2) for students with specific characteristics (i.e., gender, SES and abstract reasoning) were investigated by adding three-way interactions to the model.

Table 4. Multilevel analysis of the effects of condition (0 = control, 1 = experimental), time (1 = time 1, 2 = time 2, 3 = time 3), study track (1 = science and mathematics, 2 = engineering, 3 = Latin and mathematics), gender (1 = male, 2 = female), abstract reasoning, and SES on cognitive outcomes

Fixed effects	Phys. know.		Phys. app.		Math. know.		Math. app.		Techn.		IPM	
	β	SE	β	SE	β	SE	β	SE	β	SE	β	SE
Intercept	-0.01	0.08	0.07	0.09	-0.20	0.13	0.03	0.12	-0.06	0.08	0.02	0.08
Main effects												
Condition [0]	-0.05	0.05	-0.06	0.06	-0.09	0.12	**-0.20***	**0.09**	**-0.15****	**0.05**	-0.10	0.05
Time [1]	0.02	0.03	-0.02	0.03	-0.01	0.04	-0.07	0.04	**-0.09****	**0.03**		
Time [2]			-0.05	0.04	-0.08	0.04	-0.06	0.04	-0.06	0.03	-0.03	0.03
Study track [1]	0.00	0.08	-0.15	0.09	0.17	0.11	0.07	0.11	**0.15***	**0.08**	0.03	0.08
Study track [2]	0.03	0.09	**-0.19***	**0.10**	0.27	0.15	0.19	0.13	0.12	0.08	0.15	0.09
Gender [1]	0.02	0.04	**0.18*****	**0.04**	0.09	0.05	-0.01	0.05	0.03	0.04	-0.02	0.04
Abstract reasoning	0.01	0.02	**0.09*****	**0.02**	**0.08****	**0.03**	-0.00*	0.03	-0.03	0.02	0.03	0.02
SES	0.02	0.02	**0.04***	**0.02**	0.04	0.02	0.04	0.02	-0.00	0.02	0.00	0.02
Two-way interaction												
Condition [0]×time [1]	0.20	0.16	-0.11	0.15	**0.65*****	**0.19**	0.23	0.19	**0.30***	**0.14**		
Condition [0]×time [2]			-0.08	0.15	**0.64****	**0.19**	**0.74*****	**0.20**	**0.45****	**0.16**	0.21	0.13
Three-way interactions												
Con. [0]×time [1]×study track [1]	-0.31	0.18	0.31	0.17	**-0.65****	**0.21**	0.02	0.21	-0.10	0.16		
Con. [0]×time [2]×study track [1]			0.14	0.18	-0.29	0.22	**-0.57***	**0.23**	-0.30	0.19	**-0.29***	**0.15**
Con. [0]×time [1]×study track [2]	0.21	0.28	**0.78****	**0.24**	-0.56	0.31	0.05	0.31	-0.27	0.22		
Con. [0]×time [2]×study track [2]			0.29	0.25	**-0.97****	**0.32**	**-1.49*****	**0.34**	-0.13	0.26	**-0.43***	**0.20**
Con. [0]×time [1]×gender [1]	0.04	0.17	**-0.36***	**0.14**	-0.22	0.18	-0.06	0.18	0.12	0.13		
Con. [0]×time [2]×gender [1]			-0.11	0.16	0.02	0.20	0.13	0.21	-0.02	0.16	0.20	0.12
Con. [0]×time [1]×abs. reas.	0.08	0.08	-0.03	0.07	0.14	0.08	**0.22***	**0.08**	0.01	0.06		
Con. [0]×time [2]×abs. reas.			-0.03	0.08	**0.20***	**0.09**	0.09	0.09	0.03	0.07	0.01	0.06
Con. [0]×time [1]×SES	-0.06	0.06	**-0.13***	**0.06**	-0.09	0.08	-0.04	0.08	0.02	0.06		
Con. [0]×time [2]×SES			-0.06	0.08	0.09	0.10	-0.01	0.10	-0.02	0.08	0.01	0.06
Random effects												
ICC student	**.02***		**.13*****		**0.23****		0.09		.01		.11	
ICC school	**.02****		.02		**0.16*****		**0.07****		.01		**.04***	

Note. *p<.05; **p<.01; ***p<.001. Non-reference categories are specified between brackets.

RESULTS

Mixed models were constructed for each cognitive outcome, containing the main effects of condition, time, study, gender, abstract reasoning ability, and SES, the interaction effect of condition x time, and three-way interactions of condition x time with the other predictors. The results, including intra-cluster correlation coefficients (ICC) of the two levels (students and school) are summarized in **Table 4**. Employing dummy coding, the last category was used as a reference category each time.

Inspection of the ICC indicated that the correlation between scores of the same student (that were not explained by the model) was 2% for physics knowledge, 13% for physics application, and 23% for mathematics knowledge. Correlation between schools was 2% for physics knowledge, 16% for mathematics knowledge, 7% for mathematics application, and 4% for integrated physics and mathematics. Note that 'student ID' is nested in 'school', so that the correlations of scores with regard to the same student entail correlations within the same school. For instance, no extra ICC for students was found for physics knowledge (2%), as the ICC of school was already 2%.

General Intervention Effects

We examined to what extent cognitive performances in terms of physics (knowledge and application), mathematics (knowledge and application), technological concepts, and integrated physics and mathematics questions are impacted by the iSTEM intervention (research question 1). More specifically, we investigated whether or not students in the experimental schools would perform better on STEM concepts than students in the control schools. Additionally, we also investigated whether or not students in the experimental condition would perform better after two years than after one year, in comparison to the control group, by examining the scores in the two different conditions over time.

The univariate analyses for the six cognitive outcomes are presented together in **Table 4**. The interaction between condition and time, indicating the effect of the iSTEM intervention, is displayed underneath the 'two-way interaction' header. This interaction was significant for mathematics (knowledge and application), and
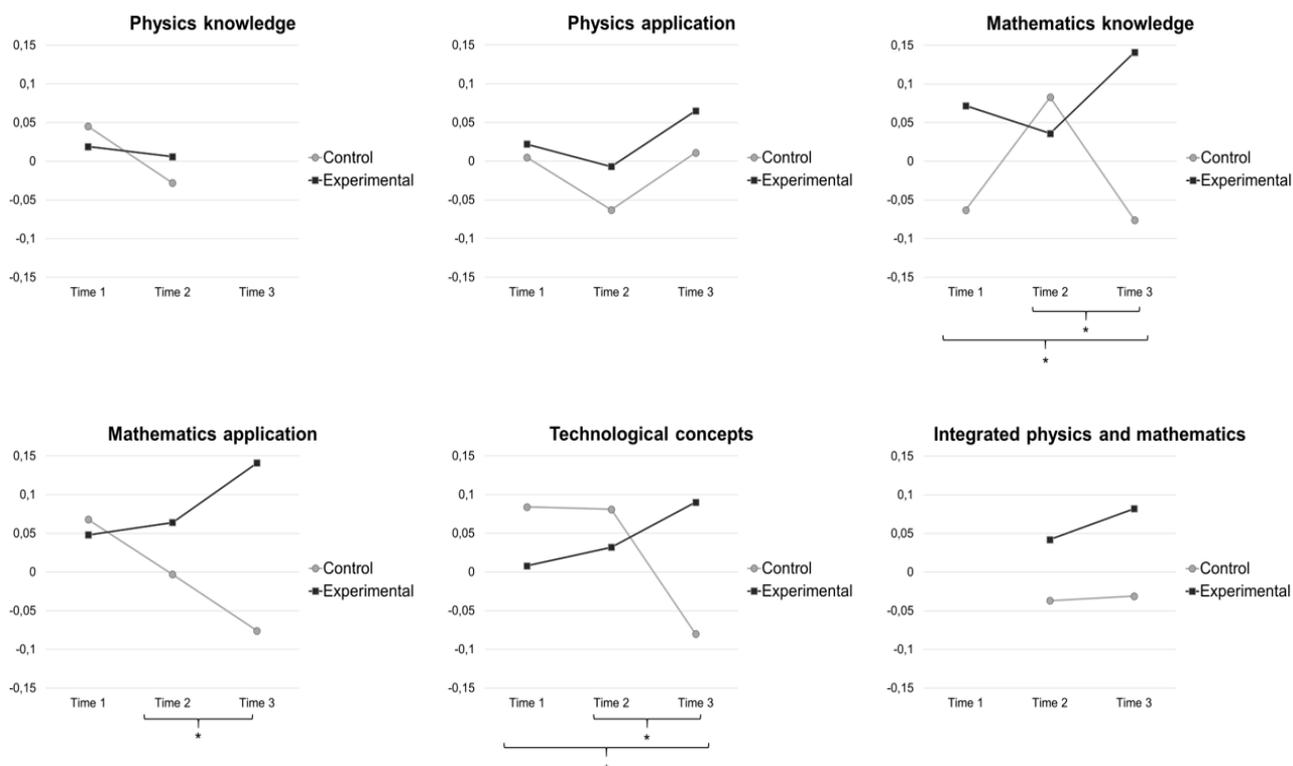
Figure 2. IRT scores on physics (knowledge and application), mathematics (knowledge and application), technological concepts and integrated physics and mathematics in control and experimental conditions on pre-test (= Time 1), post-test 1 (= Time 2), and post-test 2 (= Time 3).

Note. Significant interactions between condition and time are indicated with an asterisk.

technological concepts. No significant interaction was found for physics (knowledge and application) and integrated physics and mathematics. This finding indicates that iSTEM education mainly has an effect on cognitive performances in terms of mathematics and technological concepts.

A closer inspection of the interaction effect between condition and time for all cognitive outcomes can be found in **Figure 2**. The scores on the six outcomes are graphically displayed for control and experimental conditions across the three measurement moments. Note that these are IRT scores at a particular time-point, which means that this gives information about the relative scores of students on this time-point, but not about general progress over time.

In the case of physics knowledge, physics application and integrated physics and mathematics, no significant differences were found between the control and experimental condition over time. For mathematics knowledge, mathematics application and technological concepts, significant interactions were found. After two years, students in the experimental condition scored significantly higher on mathematics knowledge than did students in the control condition. The same result was found for technological concepts. However, both for mathematics knowledge and technological concepts, no significant difference between conditions could be found after one year. This finding indicates that the difference between the control and the experimental condition would become more pronounced after two years of iSTEM. For three of the six outcomes (mathematics knowledge, mathematics application and technological concepts), a difference was found between the scores after the first year compared to the scores after the second year. In addition, from this perspective, students in the experimental condition performed better than students in the control condition.

Differential Intervention Effects

Besides the general intervention effects, we also examined the differential cognitive effects of an iSTEM curriculum with regard to students with specific characteristics (research question 2). More specifically, we investigated whether the effects of the iSTEM intervention differed for boys or girls, students with different SES, and students with different levels of abstract reasoning ability.

The interaction between condition, time, and specific student characteristics, indicating the differential effect of the iSTEM intervention, is displayed in **Table 4** underneath the 'three-way interaction' header. The relationship between condition and time differed according to the study track for mathematics (knowledge and application)

and technological concepts. We found a remarkable result for the effect of gender on the physics application scores. In general, male students performed better on this subject than did female students. However, females in the experimental condition performed significantly better after two years than females in the control condition, while no difference was observed for males. Abstract reasoning ability might to a certain extent positively determine the scores with regard to cognitive outcomes (i.e., for physics application and mathematics knowledge), but in the case of mathematics knowledge and application, the condition determines the impact of this relationship. For students in the experimental condition, the impact of abstract reasoning ability on mathematics (knowledge and application) was larger than for students in the control condition. With regard to the SES of students, a three-way interaction between condition and time was found for physics application. The relationship between SES and scores on physics application was stronger for students in the control condition. Students with higher SES have an advantage over students with lower SES when it comes to their scores on physics application, but this advantage was more pronounced in the control condition than in the experimental condition. Otherwise stated, the impact of SES was lower for students in the experimental condition of integrated STEM.

DISCUSSION

The aim of this study was to investigate the effect of an iSTEM curriculum on students' cognitive performances regarding physics (both knowledge and application), mathematics (both knowledge and application), technological concepts, and integrated physics and mathematics. We answered two research questions: (1) what is the impact of an iSTEM curriculum on cognitive performances after one and two years, and (2) what is the differential effectiveness of an iSTEM curriculum with regard to student characteristics?

General Intervention Effects

Aligned with previous research (e.g., Becker and Park, 2011), our study highlights the potential of an iSTEM education approach on diverse cognitive outcomes. However, some differences were found regarding the domains on which the intervention of an integrated approach had an impact. Becker and Park (2011), Honey et al. (2014), and English (2016) pointed out that the positive impact of iSTEM education differed for science (i.e., physics) and mathematics, with less evidence of a positive effect on mathematics' outcomes. Our results contrasted with these findings from previous research, as we found a positive impact of iSTEM on mathematics (both for knowledge and application), but not for physics.

Students in the experimental condition scored significantly better on mathematics than students in the control condition after two years of the intervention. For students in the experimental condition, the relevance of mathematics might have become clearer and less abstract throughout the learning modules, leading to an improved understanding of mathematical concepts and applications (Fitzmaurice, O'Meara and Johnson, 2021). While outcomes on both mathematics knowledge and mathematics application could be considered as medium effects according to Cohen (1988), the largest effect was found for mathematics application. An explanation for this finding might be that students in the iSTEM condition are more familiar with applying concepts of one subdomain to another. In this way, the ability to apply STEM concepts might be facilitated.

In this study, we did not find an intervention effect in terms of physics knowledge or application. That we did not these find these effects does not necessarily suggest that such an intervention could not have benefits regarding cognitive physics' outcomes. A possible explanation for the absence of positive effects with regard to physics knowledge might be that no data were collected on the third measurement moment due to the poor psychometric qualities of the test. Consequently, we could only analyze the difference between the first and the second measurement moment with regard to the two conditions. The fact that we did not find significant differences between the two conditions with regard to these two measurement moments might not be surprising, as no significant results were found for the other cognitive outcomes of this study either. Only when the third measurement moment (i.e., after two years of iSTEM education) was taken into account, were significant differences between experimental and control condition found. Presumably, this might also be the case for the outcomes regarding physics knowledge. However, this does not explain why we did not find an effect in terms of physics application. Given the larger effects with regard to mathematics application compared with mathematics knowledge, and given the findings from previous studies (e.g., Becker and Park, 2011), we might have expected an apparent effect on physics application as well.

The contrasting findings with those of previous research (i.e., effect on physics versus mathematics) might potentially be a consequence of differences in the operationalization of the intervention. The number and the degree of integration of the different components of STEM might, for example, be decisive factors. Also, interventions could differ in their emphasis on particular concepts or topics. It has been reported by English (2016) and Yildirim (2016) that a skewed focus on science at the expense of mathematics, and no integration of all subjects,

is a common limitation within iSTEM education research. As little research exists involving the integration of all STEM components, further empirical research on the effects of iSTEM education needs to be conducted to extend the findings of the current study.

Analogous to the results with regard to mathematics knowledge, a positive effect of iSTEM education was found in terms of the results regarding technological concepts. Students in the experimental condition scored higher than students in the control condition after the third measurement moment (i.e., after two years of iSTEM education), both when compared to the first and the second measurement moments. The effect size of iSTEM on technological concepts was small (Cohen, 1988), in contrast to the effect sizes of mathematical outcomes, which were medium. This result indicates that further growth might be possible, by more explicitly addressing the technological concepts within the learning modules of the intervention. With regard to integrated physics and mathematics, no significant results were found. This result is remarkable, as the curriculum explicitly focused on the integration of the STEM domains. This finding demonstrates that it is not because connections between STEM domains are emphasized in the curriculum, students' own ability to integrate knowledge and skills necessarily improves. Thus far, the intervention of an integrated STEM curriculum appears to only effect cognitive outcomes regarding separate domains.

But the impact of an iSTEM educational approach might go beyond the cognitive outcomes measured in this study. Irrespective of the effects on the measured cognitive outcomes, the iSTEM approach could motivate students to see real-world applications and the relevance of the different STEM fields, even though students' performance did not increase in this study (Becker and Park, 2011).

To summarize, a positive impact of an iSTEM approach was found with regard to mathematics (knowledge and application) and technological concepts. Our findings indicate that the positive impact of iSTEM education is not limited to science but could also positively influence cognitive scores in other domains. From this perspective, it is important for future initiatives to explicitly incorporate all STEM domains in teaching materials, and not only select two pragmatic combinations such as physics and technology. As already mentioned, no differences were found between conditions after one year of iSTEM. This stresses the importance of a long-term iSTEM approach and has implications for the design of new integrated STEM programs. Long-term approaches with iSTEM incorporated in the standard curriculum are better suited to increase students' cognitive performance than short-term interventions.

Differential Intervention Effects

Our results showed an interesting positive impact on the performance of females with regard to 'physics application' (while no changes were found for the performances of males). As the lower physics scores of females is a well-known concern in the literature (Halpern et al., 2007), this might be an extra argument for the implementation of an integrated approach to STEM. Also, integration with more STEM domains (e.g., biology) or arts (so called 'STEAM') might have extra positive effects for girls. Previous research shows that girls tend to be more interested in biology than in physics (Baram-Tsabari and Yarden, 2008), and that STEM activities are more appealing to girls when also arts are involved (Boyle, 2021).

Scores on physics application differed for students with lower SES, when the integrated STEM condition was compared with the traditional approach. The negative impact of low SES (Yerdelen-Damar and Peşman, 2013) was smaller for students in the experimental condition of iSTEM. From this perspective, iSTEM education might create more equity.

For students in the experimental condition, the impact of abstract reasoning ability on mathematics (knowledge and application) was larger than for students in the control condition. This finding implies that, with regard to mathematics, an iSTEM approach favors those who already have more cognitive capabilities. The challenging nature of the learning modules might provide an explanation for this finding. When designing an integrated STEM intervention, educators should bear in mind that the impact of the intervention could vary with reasoning ability.

Limitations and Directions for Future Research

While our study has several strengths, such as its scale and longitudinal design, the explicit focus on iSTEM (the inclusion of all STEM domains), and the inclusion of multiple cognitive outcomes, limitations should also be acknowledged. First, we compared experimental schools with control schools, but it could not be guaranteed that the experimental schools implemented the iSTEM intervention in an impeccable way (O'Donnell, 2008), and that students in the control condition had no STEM initiatives whatsoever in their schools. It is more plausible that the experimental schools varied in the extent to which they implemented the intervention as intended, and that the control schools varied in the degree to which they did not implement (other) STEM initiatives. Nevertheless, we could ensure that all experimental schools were familiar with the integrated approach and that they implemented the learning modules in their classes. Also, intervisions with experimental schools and educational umbrellas were

regularly organized so that schools were guided through the process and could optimize their iSTEM approach. The control schools, on the other hand, were queried about their ongoing STEM initiatives during the study. Most control schools did provide STEM projects for their students. However, they were only small-scale (e.g., extra programming exercises) and did not follow an integrated approach. In conclusion, we could presume that the critical component of the intervention (i.e., an iSTEM approach) was not present in the control condition. Another component in which experimental and control schools could differ was the time-on-task. While iSTEM was taught in the regular hours for mathematics and physics, schools still had the autonomy to allocate optional hours to the iSTEM learning modules. This extra time-on-task could accordingly have led to improved cognitive performances on STEM domains. But control schools might also have optional hours in which students can opt for further deepening in mathematics or sciences. And even when time-on-task was different for experimental and control schools, we know from earlier research that variation in cognitive performance is not consistently explained by differences in instructional time: only 1 to 15 percent of the variance was explained by time-on-task (Karweit, 1984), and explained variance varied depending on the time-on-task estimation method (Kovanovic et al., 2015). Therefore, providing extra time is not a sufficient condition for learning to take place. Also, if time-on-task was a crucial factor, we could expect that all learning outcomes would benefit from extra learning time. However, this was not the case in the current study. Future research could measure different characteristics (e.g., degree of integration, presence of a design challenge, time-on-task, etc.) of STEM initiatives in experimental and control settings and determine the relationship with students' cognitive outcomes. In this way, a measure for implementation fidelity in the experimental setting could be provided, and a stricter oversight of the control condition could be attained.

A second limitation is that the role of the teacher was not included explicitly in this study. Variations in the implementation of the intervention could be mainly attributed to teacher characteristics and practices. Factors such as teachers' individual characteristics when accepting a new instructional approach, their attitudes towards an integrated approach with regard to STEM education, and their prior experiences could have an influence on the implementation of the learning modules (Thibaut et al., 2018b; Henderson and Dancy, 2007). Although teacher influence could be partially accounted for by controlling for the random effect of schools, we suggest that future research incorporates teacher characteristics when investigating the effect of iSTEM education. At the same time, further research is needed on ways of assisting teachers to implement iSTEM learning modules (Moore and Smith, 2014; Sevimli and Ünal, 2022).

Summarizing, research regarding educational interventions is complex due to its multifaceted nature, such as the impact of implementation fidelity, teachers' characteristics, and complex settings. This exploratory study provided a first insight into the effects of iSTEM on a wide range of cognitive outcomes and encourages future research to further investigate the crucial ingredients and the effective mechanisms associated with an iSTEM education.

CONCLUSION

This longitudinal study revealed that iSTEM education had positive effects on cognitive performance on mathematics knowledge and application, and technological concepts. Furthermore, the intervention had a positive impact on the performance of females on physics application, the negative impact of low SES was smaller in the case of physics application, and students with high abstract reasoning capabilities were favored when it came to mathematics knowledge and application. Our research shows the importance of a long-term integrated STEM educational approach and advocates the integration of all STEM domains in educational initiatives.

ACKNOWLEDGEMENT

REFERENCES

Baram-Tsabari, A. and Yarden, A. (2008). Girls' biology, boys' physics: Evidence from free-choice science learning settings. *Research in Science & Technological Education*, 26(1), 75-92. https://doi.org/10.1080/02635140701847538

Becker, K. and Park, K. (2011). Effects of integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A preliminary meta-analysis. *Journal of STEM Education*, 12(5/6), 23-27.
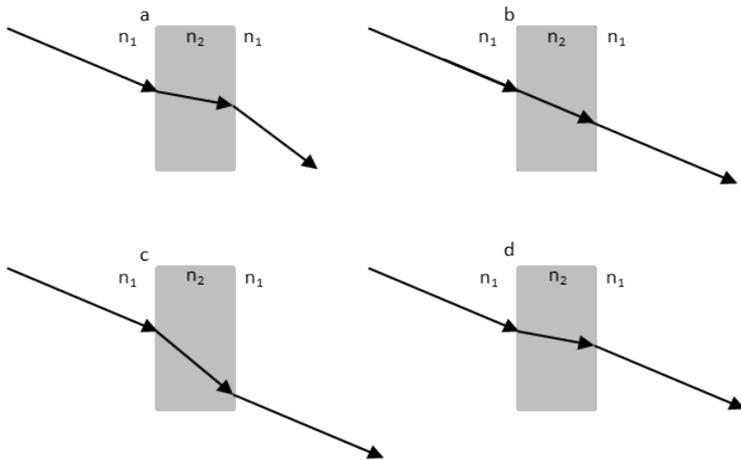
Bøe, M. V., Henriksen, E. K., Lyons, T. and Schreiner, C. (2011). Participation in science and technology: Young people's achievement-related choices in late-modern societies. *Studies in Science Education*, 47(1), 37-72. https://doi.org/10.1080/03057267.2011.549621

Bornstein, M. H. and Bradley, R. H. (2003.) *Socioeconomic status, parenting, and child development*. Mahwah, NJ: Lawrence Erlbaum Associates.

Boyle, J. (2021). Oceans of inspiration: A marine based STEAM project. *European Journal of STEM Education*, 6(1), 15-36. https://doi.org/10.20897/ejsteme/11356

Bradley, R. H. and Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53(1), 371-399. https://doi.org/10.1146/annurev.psych.53.100901.135233

Cantrell, P., Pekca, G. and Ahmad, I. (2006). The effects of engineering modules on student learning in middle school science classrooms. *Journal of Engineering Education*, 95(4), 301-309. https://doi.org/10.1002/j.2168-9830.2006.tb00905.x

Cohen, J. (1988). *Statistical power analysis for the behavioral science (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associations.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J. and Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30(2), 163-183. https://doi.org/10.1016/S0160-2896(01)00096-4

Deary, I. J., Strand, S., Smith, P. and Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35(1), 13-21. https://doi.org/10.1016/j.intell.2006.02.001

De Loof, H., Boeve-de Pauw, J. and Van Petegem, P. (2022). Engaging Students with Integrated STEM Education: a Happy Marriage or a Failed Engagement?. *International Journal of Science and Mathematics Education*, 20(7), 1291-1313. https://doi.org/10.1007/s10763-021-10159-0

English, L. D. (2016). STEM education K-12: Perspectives on integration. *International Journal of STEM Education*, 3(1), 3. https://doi.org/10.1186/s40594-016-0036-1

Fitzmaurice, O., O'Meara, N. and Johnson, P. (2021). Highlighting the relevance of mathematics to secondary school students--Why and how. *European Journal of STEM Education*, 6(1), 7-19. https://doi.org/10.20897/ejsteme/10895

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S. and Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 8(1), 1-51. https://doi.org/10.1111/j.1529-1006.2007.00032.x

Henderson, C. and Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Physics Education Research*, 3(2), 1-14. https://doi.org/10.1103/PhysRevSTPER.3.020102

Hermans, S., Gijsen, M., Mombaers, T. and Van Petegem, P. (2022). Gendered patterns in students' motivation profiles regarding iSTEM and STEM test scores: A cluster analysis. *International Journal of STEM Education*, 9(1), 1-17. https://doi.org/10.1186/s40594-022-00379-3

Hernandez, P. R., Bodin, R., Elliott, J. W., Ibrahim, B., Rambo-Hernandez, K. E., Chen, T. W. and de Miranda, M. A. (2014). Connecting the STEM dots: Measuring the effect of an integrated engineering design intervention. *International Journal of Technology and Design Education*, 24(1), 107-120. https://doi.org/10.1007/s10798-013-9241-0

Honey, M., Pearson, G. and Schweingruber, H. (2014). *STEM integration K-12 education: Status, prospects, and an agenda for research*. Washington, DC: National Academies Press.

Karweit, N. (1984). Time-on-task reconsidered: Synthesis of research on time and learning. *Educational Leadership*, 41(8), 32-35.

Keith, K. (2018). *Case study: Exploring the implementation of an integrated STEM curriculum program in elementary first grade classes* [Doctoral dissertation, Concordia University, Portland].

Kjærnsli, M. and Lie, S. (2011). Students' preference for science careers: International comparisons based on PISA 2006. *International Journal of Science Education*, 33(1), 121-144. https://doi.org/10.1080/09500693.2010.518642

Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S. and Baker, R. (2015). Does time-on-task estimation matter? Implications on validity of learning analytics findings. *Journal of Learning Analytics*, 2(3), 81-110. https://doi.org/10.18608/jla.2015.23.6

Lam, P. C., Doverspike, D., Zhao, J., Zhe, J. and Menzemer, C. (2008). An evaluation of a STEM program for middle school students on learning disability related IEPs. *Journal of STEM Education: Innovations and Research*, 9(1/2), 21-29.

Lederman, N. G. and Niess, M. L. (1997). Integrated, interdisciplinary, or thematic instruction? Is this a question or is it questionable semantics? *School Science and Mathematics*, 97(2), 57-58. https://doi.org/10.1111/j.1949-8594.1997.tb17342.x

Moore, T. J. and Smith, K. A. (2014). Advancing the state of the art of STEM integration. *Journal of STEM Education*, 15(1), 5-10.

National Academies of Science. (2007). *Rising above the gathering storm.* The Committee on Prospering in the Global Economy of the 21st Century. Washington, DC: The National Academies Press.

Newton, K. J., Leonard, J., Buss, A., Wright, C. G. and Barnes-Johnson, J. (2020). Informal STEM: Learning with robotics and game design in an urban context. *Journal of Research on Technology in Education*, 52(2), 129-147. https://doi.org/10.1080/15391523.2020.1713263

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, 78(1), 33-84. https://doi.org/10.3102/0034654307313793

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd edition).* Newbury Park, CA: SAGE.

Raven, J. C., Court, J. H. and Raven, J. (1977). *Raven's progressive matrices and vocabulary scales.* New York, NY: Psychological Corporation.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1-25. https://doi.org/10.18637/jss.v017.i05

Roehrig, G. H., Moore, T. J., Wang, H.-H. and Park, M. S. (2012). Is adding the E enough? Investigating the impact of K-12 engineering standards on the implementation of STEM integration. *School Science and Mathematics*, 112(1), 31-44. https://doi.org/10.1111/j.1949-8594.2011.00112.x

Sanders, M. (2009). STEM, STEM education, STEMmania. *Technology Teacher*, 68(4), 20-26.

Sevimli, E. and Ünal, E. (2022). Is the STEM approach useful in teaching mathematics? Evaluating the views of mathematics teachers. *European Journal of STEM Education*, 7(1), 1-11. https://doi.org/10.20897/ejsteme/11775

Standish, N., Christensen, R., Knezek, G., Kjellstrom, W. and Bredder, E. (2016). The effects of an engineering design module on student learning in a middle school science classroom. *International Journal of Learning, Teaching and Educational Research*, 15(6) 156-174.

Tate, W. F. (1997). Race-ethnicity, SES, gender, and language proficiency trends in mathematics achievement: An update. *Journal for Research in Mathematics Education*, 28(6), 652-679. https://doi.org/10.2307/749636

Thibaut, L., Ceuppens, S., De Loof, H., De Meester, J., Goovaerts, L., Struyf, A., … and Hellinckx, L. (2018a). Integrated STEM education: A systematic review of instructional practices in secondary education. *European Journal of STEM Education*, 3(1), 1-12. https://doi.org/10.20897/ejsteme/85525

Thibaut, L., Knipprath, H., Dehaene, W. and Depaepe, F. (2018b). The influence of teachers' attitudes and school context on instructional practices in integrated STEM education. *Teaching and Teacher Education*, 71, 190-205. https://doi.org/10.1016/j.tate.2017.12.014

Vasquez, J., Sneider, C. and Comer, M. (2013). *STEM lesson essentials, grades 3-8: Integrating science, technology, engineering, and mathematics.* Portsmouth, NH: Heinemann.

Wang, H. H., Moore, T. J., Roehrig, G. H. and Park, M. S. (2011). STEM integration: Teacher perceptions and practice. *Journal of Pre-College Engineering Education Research*, 1(2), 1-13.

Williams, P. J. (2011). Engineering: Good for technology education. In M. J. de Vries (ed), *Position technology education in the curriculum: International technology education series* (pp. 87-99). New York, NY: Springer. https://doi.org/10.1007/978-94-6091-675-5_8

World Economic Forum. (2016). *The future of jobs employment, skills and workforce strategy for the fourth industrial revolution. Global challenge insight report.* Geneva: World Economic Forum.

World Economic Forum. (2020). *The future of jobs report 2020.* Geneva: World Economic Forum.

Yang, D. and Baldwin, S. J. (2020). Using technology to support student learning in an integrated STEM learning environment. *International Journal of Technology in Education and Science*, 4(1), 1-11. https://doi.org/10.46328/ijtes.v4i1.22

Yang, D., Baek, Y., Ching, Y. H., Swanson, S., Chittoori, B. and Wang, S. (2021). Infusing computational thinking in an integrated STEM curriculum: User reactions and lessons learned. *European Journal of STEM Education*, 6(1), 4-18. https://doi.org/10.20897/ejsteme/9560

Yerdelen-Damar, S. and Peşman, H. (2013). Relations of gender and socioeconomic status to physics through metacognition and self-efficacy. *The Journal of Educational Research*, 106(4), 280-289. https://doi.org/10.1080/00220671.2012.692729

Yildirim, B. (2016). An analyses and meta-synthesis of research on STEM education. *Journal of Education and Practice*, 7(34), 23-33.

APPENDIX A

Physics Knowledge

A light beam passes through a plate. Which of the images below is correct when the refractive indices are n1<n2?

A) a
B) b
C) c
D) d

Physics Application

The spring constant of three identical massless springs is 0.200 N/cm. What is the stretching of the feathers when they are hung next to each other in order to carry a common load with a mass of 300 g?
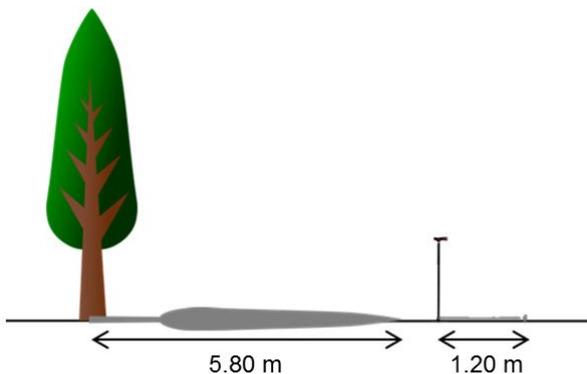A) 0.667 cm
B) 4.905 cm
C) 14.715 cm
D) 19.620 cm

Mathematics Knowledge

The directional coefficient of the line through (a, 0) and (0, b) equals:
A) a/b
B) -a/b
C) b/a
D) -b/a

Mathematics Application

Peter would like to know the height of the tree. For this purpose, he can use grandma's walking stick, which is 1.0 m long. Given the illustration below, what is the height of the tree?

A) 6.8 m
B) 5.8 m
C) 4.8 m
D) 0.2 m

Technological Concepts
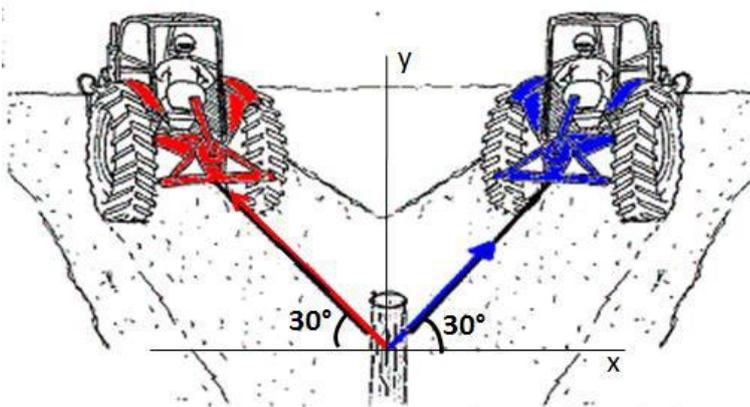
Given the program code below:
```
number i=0
number j=5
REPEAT AS LONG AS (i<3 and j+1<10) {
 PRINT (i, j)
 i=i+1
 j=j+1
 MEMORIZE i
 MEMORIZE j
 }
 PRINT (i, j)
```
what will be the printing output?
A) 0 5 1 6 2 7 3 8
B) 0 5 1 6 2 7 3 8 4 9 5 10
C) 0 5 1 6 2 7 3 8 4 9
D) 0 5 1 6 2 7 2 7

Integrated Physics and Mathematics

Two tractors are pulling a pole out of the ground. The red tractor uses a force of 5,000 N and the blue tractor uses a force of 2,500 N. What is the magnitude (in integers) of the total force that is applied on the pole in the direction of the y-axis?
A) 2500 N
B) 7500 N
C) 3590 N
D) 5590 N

APPENDIX B

The ltm-package of R (open source software for statistical computing) was employed, using latent trait models under IRT, which is fit for an analysis of multivariate dichotomous data (Rizopoulos, 2006). Difficulty (i.e., the ability required to guarantee a 50% probability of answering the item correctly) and discrimination of the items (i.e., an index of an item's capability to differentiate between students in different positions on the latent ability) were analyzed, and items with a discrimination value of less than 0.15 were removed from the item battery. Subsequently, IRT was re-performed with the remaining items. Thereafter, the model with the best fit for the data was identified by analysis of variance (ANOVA). The Rash model (i.e., all items have a discrimination index of 1 logit) was compared with the one-parameter logistic model (1-PL, i.e., the discrimination index are the same for all items, but can have a value other than 1) and with the two-parameter logistic model (2-PL, i.e., the discrimination index can vary over items). For each instrument, the model with the best fit, the initial number of items, the remaining number of items, and information regarding discrimination values (a) and difficulty (β) are presented.

Instruments Measurement Moment 1

In **Table B1**, the results of the IRT analyses of the pretest instruments (measurement moment 1) are shown. Analysis of variance (ANOVA) showed that for physics (knowledge and application) and mathematics (knowledge and application) the 2-PL model had the best fit for the data. The 1-PL model had the best fit for the data on technological concepts. All items from the mathematics tests (knowledge and application) were retained, as no item had low discrimination values ($a<0.15$). For the other instruments, one or more items were omitted.

Table B1. IRT analyses of instruments measurement moment 1

	Phys. know.	Phys. app.	Math. know.	Math. app.	Techn.	IPM
Model	2-PL	2-PL	2-PL	2-PL	1-PL	
# initial items	10	15	8	25	25	
# remaining items	9	14	8	25	18	
Min a	0.35	0.36	0.41	0.15	0.61	
Max a	1.33	1.72	1.24	1.62	0.61	/
Mean a	0.65	0.70	0.84	0.93	0.61	
Min β	-4.26	-1.94	-2.81	-2.00	-6.36	
Max β	-0.77	3.93	0.49	3.45	2.93	
Mean β	-1.88	0.30	-0.74	-0.67	2.93	

Instruments Measurement Moment 2

ANOVA showed that for all the instruments used in the first posttest (measurement moment 2) the 2-PL model had the best fit for the data (**Table B2**). For each instrument one or more items were omitted due to low discrimination values.

Table B2. IRT analyses of instruments measurement moment 2

	Phys. know.	Phys. app.	Math. know.	Math. app.	Techn.	IPM
Model	2-PL	2-PL	2-PL	2-PL	1-PL	2-PL
# initial items	29	22	10	31	25	14
# remaining items	22	15	9	29	20	9
Min a	0.19	0.20	0.24	0.29	0.18	0.18
Max a	1.87	2.02	1.37	2.38	2.31	17.91
Mean a	0.78	0.68	0.60	0.91	0.90	2.30
Min β	-2.82	-1.56	-2.30	-1.55	-1.47	-0.92
Max β	9.68	4.92	1.49	3.45	8.36	5.60
Mean β	0.73	1.09	-0.02	0.40	0.65	2.02

Instruments Measurement Moment 3

Table B3 shows the results of the IRT analyses of the second posttest (measurement moment 3). The 2-PL model best fitted the data of physics (application) and mathematics (knowledge and application), whereas the 1-PL model best fitted the data of technological concepts and integrated physics and mathematics questions. Of all 17 items in the physics knowledge test, only one item had a discrimination index of $a>0.15$, which was insufficient to perform further analysis. As a result, no reliable indicators of physics knowledge were collected in the second post-test of the study.

Table B3. IRT analyses of instruments measurement moment 3

	Phys. know.	Phys. app.	Math. know.	Math. app.	Techn.	IPM
Model		2-PL	2-PL	2-PL	1-PL	1-PL
# initial items	17	13	15	29	19	12
# remaining items	1	11	15	28	10	9
Min a		0.16	0.37	0.35	0.64	0.5
Max a		1.95	2.51	8.36	0.64	0.5
Mean a		0.64	1.45	1.77	0.64	0.5
Min β		-3.03	-1.73	-1.48	-1.07	0.83
Max β		8.39	3.76	5.06	1.90	3.54
Mean β		1.12	0.55	0.45	0.22	1.92