

Interdisciplinary Method for Assessing Students' Ability Based on STEM Projects*

QIN NI¹, LELE ZHANG¹, BO ZHANG¹ and FENG-KUANG CHIANG^{2**}

¹ College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China.

E-mail: niqin@shnu.edu.cn, claire_zll@163.com

² Department of Educational Technology, College of Education, Shanghai Normal University, Shanghai 200234, China.

E-mail: fkchiang@shnu.edu.cn

The labor markets of the information age have an urgent demand for engineering and technical workers. STEM education has evolved into a metadiscipline, which involves an integrated effort to eliminate the traditional barriers between different subjects. Recent studies have reported increased levels of STEM education in international science education at the preliminary education stage worldwide. However, STEM assessment studies have lacked systematic and comprehensive assessment standards. Therefore, this study included a series of pretest and posttest evaluation questionnaires. With multiple evaluation methods, this study analyzed the effects of the STEM courses. This paper assesses the effect of the STEM courses on a total of 693 students of a primary school in Beijing. The results can be used to establish a more systematic and perfect evaluation system for students' interdisciplinary learning ability. STEM courses have achieved certain positive effects on emotional attitudes toward STEM, subject cognition, engineering professional cognition and engineering design ability among the three grade students, particularly for middle-school students.

Keywords: STEM education; engineering design; interdisciplinary learning; ability assessment

1. Introduction

STEM (originally derived in the United States) refers to the abbreviation of “science, technology, engineering and mathematics.” STEM curriculum focuses on strengthening students' education in four areas: (1) scientific literacy, involving the use of scientific knowledge (e.g., that of physics, chemistry, biological sciences, and geospatial sciences) to understand nature and participate in processes that affect nature; (2) technical literacy, the ability to use, manage, understand and evaluate technology; (3) engineering literacy, the understanding of the technical processes of engineering design and development; and (4) mathematical literacy, the ability of students to discover, express, explain and solve mathematical problems in various situations. Over time, STEM education has received considerable attention; many scholars have published studies on STEM education. Furthermore, many countries have attached great importance to STEM education and have regarded STEM education as a focal point for enhancing national competitiveness in science and technology and for training creative people [1, 2]. Many studies on STEM education only focus on the designing of STEM courses but lack any consideration of systematic evaluation standards in China. Most works lack systematic assessment

methods; therefore, establishing a comprehensive evaluation system to fully consider the STEM curriculum results is necessary.

2. Related work

STEM education researchers should pay attention to not only the integration of different disciplines but also the effect of these courses and activities on students in terms of cognition regarding knowledge, attitudes toward things, and the ability to think. At present, relevant studies have been performed in the United States, Australia, the United Kingdom, and some Asian countries; among these studies, the studies from the United States are relatively detailed. Numerous scholars are still exploring the best approach to study by analyzing crossdisciplinary programs by using diverse tools. Some scholars have directly used common or international assessment, such as the Texas Assessment of Knowledge and Skills test and Grade Level Content Expectations [3, 4]. Some studies have integrated different subject tests based on national standards to observe students' changes on certain subjects [5–8]. For instance, Parker et al. [8] used NEAP to argue that instructional approaches based on engineering design process could meet STEM needs more effectively than earlier approaches. The problem with this evaluation method is that it may not correspond to the teaching goals well, so we suggest using different evaluation methods targeted at certain indexes.

** Corresponding author.

These authors contributed equally to this work and should be considered co-first authors.

For assessing students' attitude toward interest, career tendency, utilization of some skills, including self-efficacy, self-making scales were mostly frequently used; assessment often includes numerous dimensions. For instance, Peterman et al. [7] used Career Interest Questionnaire (CIQ) to effectively measure the career attitudes of middle and high school after technology-enhanced STEM experiences; Gilliam et al. [9] and Guzey et al. [10] used integrated questionnaires based on research goals and found that students had significant improvement in several abilities: cognitive ability, problem solving and creativity. One classic evaluation scale was the STEM semantic survey, which compared students' attitudes toward subjects and careers. Students were asked to choose one of two opposite adjectives, such as interesting and boring, exciting and dull [11]. In addition, some other scales, such as the Likert scale, use descriptive sentences. To assess student abilities, scales were also utilized. Some other studies used rubrics marked by teachers to test students' abilities [12]. Fan et al. [13] developed a design project rubric including several dimensions with four levels, and proved that students benefited from higher-order thinking skills and engineering design abilities in the STEM program.

Because past STEM courses were often guided by engineering activities, most studies considered engineering design ability the most essential. Setting a scenario task was often used in project evaluations that involved asking students to propose a solution and produce a prototype. Wilson et al. [14] asked students to complete tasks for designing furniture and water systems on a campus, and investigated students' abilities by observing and coding their behavior according to procedures of engineering design. Although these methods may be somewhat complex, they could test students' problem solving abilities in situations resembling real life and determine the students' weak skills for more targeted instruction. However, these studies also had problems. For instance, some scales lacked examination of reliability, validity and in-depth quantitative analysis. Some rubrics lacked detailed scoring criteria, which may have caused substantial subjective errors [15]. Other qualitative evaluation methods included interviews and observation of students' drawings. Sometimes these methods are more suitable for young students because researchers can examine the inner thoughts of students in detail. For instance, Lyons et al. [16] found that students tend to define engineers as manual workers engaged in engineering activities, but males are not quite clear about what engineers do.

In addition, La Trobe University developed an activity, LaserTag, the effectiveness of which was measured based on anonymous student surveys

evaluating students pretest and posttest interests in engineering and the STEM disciplines. The survey showed that 97% of the participants believed that the event was "highly enjoyable" or "enjoyable" and that 55% of students who were previously unsure about engineering as a career "strongly agreed" or "agreed" that they were more interested in studying engineering after participation [17]. The evaluation level of the survey results is too singular, and the grade evaluation description is not specific enough; therefore that study cannot be systematically combined with other studies. Numerous diversified evaluation methods pay considerable attention to process evaluation including scale evaluation, situational evaluation, student reflection diary analysis, drawing evaluation, classroom observation, afterschool interviews and presentation of results. Some STEM activities focus on improving students' levels. They evaluate students' learning effects by directly use mathematics scores, related subject test questions [18, 19, 20], and international scientific tests, such as PISA, TIMSS and NEAP. Although such an evaluation study can analyze the effect of different variables on educational outcomes, the form is too simple and has certain limitations.

Scale evaluation is the method most widely used by researchers to evaluate the emotional attitudes of students in STEM education. It is used to assess students' attitudes toward science, technology, engineering, mathematics and STEM occupations. Students are asked to choose a level that matches their characteristics from two adjectives. When evaluating the expected changes in the MSOSW ITEST project activity, Tylerwood et al. [11] created two tools for analysis, applying personal scales to the STEM Semantic Survey and the STEM career interest survey. In addition, it was found that they had excellent internal consistency and reliability. In an educational study, key elements of the study included sampling techniques, ethical considerations, data collection methods, measurements, judgment validity, experimental and nonexperimental methods, description and reasoning statistics, qualitative data analysis, and report preparation [21]. In numerous STEM papers involved in vocational education, researchers have also used declarative statement scales to evaluate students' career orientation toward STEM [22].

Engineering design capability is a key indicator in STEM education. Given the clear importance of high-quality STEM education and the limitations of the current education system for achieving high-quality STEM education goals, Emily Saxton and others designed an educational evaluation method to improve the theoretical basis and systems thinking for the improvement of STEM education; they

proposed a universal measurement system for K-12 STEM education. These changes in the assessment and research focus supported by the common measurement system can greatly facilitate the transformation of STEM education [23].

The current works on STEM assessment exhibit a lack of detailed analysis of the gender factors. However, many countries attach importance to the development of STEM education, especially to encourage more women to join the STEM field [24]. For instance, the United Arab Emirates has attracted more women than men to STEM fields. The education system of the United Arab Emirates has encouraged numerous Emirati women to graduate with degrees in STEM fields, especially in the aerospace sector [25]. Richardson used quantitative and qualitative methods to study gender differences in higher education experiences and developed simple quantitative indicators. Examination of the questionnaire reply form reveals obvious differences between male and female students in the practices of education and teaching [26]. In a survey based on the attitude toward engineers and scientists of middle-school students, it is found that female students believe that scientists can make money and work alone but male students more frequently connect these indicators with engineers [27]. When researchers use the mapping analysis method, they designed a detailed coding analysis system, which was divided into four dimensions: engineering products, engineering processes, engineering fields and engineering portraits. Studies have shown that before the start of the STEM course, students were more inclined to define engineers as manual workers. Most of the engineers were drawn as males and the students' understanding was relatively one-sided. After the end of the STEM course, students' knowledge of engineers was more abundant and deeper. Furthermore, the proportion of female engineers was significantly increased [28]. These studies indicate that women's role in STEM education is obvious but does not involve gender-based assessment criteria.

In general, typical existing STEM evaluation studies still have a single evaluation dimension but lack systematic and comprehensive assessment methods. Therefore, this paper is based on an STEM project and combines multiple methods for evaluating students' interdisciplinary learning ability to establish a more systematic and faultless evaluation system.

3. Methodology

3.1 Study design

In total, 693 students from the grades 2, 4 and 6 of a Beijing primary school were invited to participate in

a STEM course. In this experiment, standard single-group pretest and posttest were adopted. Students from three grades participated in three different STEM courses; these courses lasted 8 weeks and had two lessons per week. The pretest and posttest were conducted in the first and last week of the course respectively, lasting for 40 minutes. All teachers were armed with same teaching resources to guarantee that the teaching quality was consistent.

3.2 Course introduction

This STEM program covered many topics, including knowledge of science, technology, engineering, and mathematics. The course was mainly guided by engineering philosophy and followed a specific sequence of tasks: project introduction, project implementation, outreach activities and project evaluation. In the course, students learned certain subject knowledge, understood the occupation of engineers, performed experimental exploration and participated in other group activities. Finally, they were asked to design and produce engineering artifacts and then to present, evaluate and reflect on them.

3.3 Tools and data analysis

First, questionnaires for pretest and posttest were developed; these were used to evaluate students' learning attitudes and achievements based on research questions and course context. Each questionnaire consisted of four parts: personal information, attitude scale, subject cognition test and engineering design task. The first part contained surveys of each student's interest in the subjects: mathematics, Chinese, English, sports, morality, computer, art and science. The second part consisted of 12 items, which could be divided into the four following dimensions; each dimension had three items. All items referred to previously published studies and standard scales: Wendell [29], Williams Scale [30], Baylor [31], Ragusa [32], Torrance Test of Creative Thinking [33] and Chinese Version of Creative Disposition Inventory (CTDI-CV) [34].

The attitude scale part of questionnaires for grade 2 retained eight items, which were divided into four dimensions: cooperative communication, problem solving, hands-on practice and engineer career orientation. Each dimension had two items and the Cronbach's α value was 0.684. Questionnaires for grades 4 and 6 retained 12 items, which were divided into four dimensions: cooperative communication, problem solving, critical thinking, and engineer career orientation. Each dimension had three items and the Cronbach's α values were 0.741

and 0.783 for grades 4 and 6 respectively; these values obtained high reliability.

The third part of the test was designed according to their textbook, covering the following key points, science, technology, engineering and mathematics, with the proportions of the key points were respectively 55%, 11%, 20% and 14% in grade 2, 50%, 10%, 30% and 10% in grade 4, 50%, 37.5% and 12.5% in grade 6. The fourth part was intended to test students' engineering design abilities. They were required to design an artifact or device in a certain scenario (closely related with the course topic), to draw a sketch and to offer an explanation. The evaluation of this part was divided into three dimensions: problem and demand definition, problem solving ability and designing ability with three or four levels for each dimension. To determine the validity of the scores, three evaluators (graduate students majoring in science and technology) scored the work samples of three classes (each class was randomly selected from each grade).

Statistical Package for the Social Science was used to analyze the data and the correlation analysis was performed with statistical methods, such as the Pearson's correlation coefficient, Kendall correlation coefficient and Student's t test. Learning attitudes were further divided into six dimensions, namely collaborative communication, creativity, critical thinking, hands-on practice, problem solving and professional cognition. The paper adopts forms of the standard evaluation scale [29], Williams Creativity Scale [30] and Adolescent Scientific Creativity Scale [35]. The emotional attitudes of

grades 4 and 6 are in the form of a 5-point scale. Considering that the second-year students have limited cognitive ability, some adjustments were made to popularize language expression and to communicate with grade 2 students in the form of a 3-point scale.

4. Results and discussion

4.1 Degree of affection

Due to the different preferences of each person, students of different grades have different degrees of affection for each subject. Grades 2, 4 and 6 were divided for three periods (September 2016, December 2016 and June 2017) to investigate the changes in students' preferences for each subject.

4.1.1 Grade 2

Table 1 lists the three test changes in the degree of affection of grade 2 students for different subjects during a school year. Here, “***” is significantly correlated at the 0.01 level (both sides) and “**” is significantly correlated at the 0.05 level (both sides). It can be seen that the degree of students' affection for various subjects did not change much in the three tests. The high degrees of affection were mainly concentrated in three disciplines of science, art, and sports. The subject with the lowest degree of affection was math. In the two posttests, the students' affection for math significantly improved (both $p < 0.05$). Compared with the test in September 2016, students' attitudes toward moral education significantly improved at the posttests in the

Table 1. Three test changes to degree of affection of different disciplines in second-year students

	2016.09(T1)		2016.12(T2)		2017.06(T3)		Paired t-value (T2-T1)	Paired t-value (T3-T1)
	Mean	Std	Mean	Std	Mean	Std		
Math	3.17	1.875	3.37	1.952	3.39	1.811	1.491*	1.008*
Chinese	3.93	1.569	3.96	1.657	3.94	1.638	1.611	0.179
English	3.69	1.762	3.81	1.743	3.71	1.661	-3.028	-1.041
Sports	4.20	1.366	4.24	1.387	4.32	1.404	-3.994	1.414
Moral education	3.91	1.483	4.11	1.552	4.07	1.570	1.512*	0.287**
Art	4.32	1.308	4.49	1.151	4.44	1.249	-0.053	-0.548
Science	4.60	1.030	4.61	1.105	4.49	1.274	-0.513	0.854

Table 2. Three test changes to degree of affection of different disciplines in fourth-year students

	2016.09(T1)		2016.12(T2)		2017.06(T3)		Paired t-value (T2-T1)	Paired t-value (T3-T1)
	Mean	Std	Mean	Std	Mean	Std		
Math	3.25	0.88	3.23	0.85	3.30	0.77	1.31	0.82
Chinese	3.32	0.83	3.28	0.75	3.40	0.75	2.25*	1.72
English	3.05	0.92	3.13	0.96	3.01	1.05	-1.41	-0.38
Sports	3.42	0.92	3.32	0.98	3.20	1.04	-1.50	-2.79**
Moral education	3.49	0.72	3.56	0.78	3.59	0.69	0.36	1.66
Computer	3.70	0.59	3.82	0.53	3.79	0.58	-0.75	1.96
Art	3.57	0.79	3.72	0.57	3.65	0.79	-1.33	1.31
Science	3.76	0.46	3.78	0.48	3.30	0.89	-7.00**	-6.24**

Table 3. Three test changes to degree of affection of different disciplines in sixth-year students

	2016.09(T1)		2016.12(T2)		2017.06(T3)		Paired t-value (T2-T1)	Paired t-value (T3-T1)
	Mean	Std	Mean	Std	Mean	Std		
Math	4.16	0.92	4.33	0.99	4.11	1.09	0.20*	-0.07
Chinese	3.93	1.08	4.07	1.19	3.72	1.25	0.21*	-0.23
English	3.84	1.20	4.03	1.21	3.65	1.30	0.25	-0.12
Sports	4.03	1.18	4.08	1.27	4.03	1.27	0.10	0.06
Moral education	4.19	1.10	4.31	1.06	4.11	1.18	0.18	-0.10
Computer	4.14	1.25	4.17	1.23	3.98	1.29	0.06	-0.10
Art	4.32	1.11	4.34	1.02	4.18	1.17	0.06	-0.16
Science	4.28	1.05	4.37	1.17	4.03	1.27	0.14	-0.18

first and second semesters ($p < 0.05$ and $p < 0.01$, respectively).

4.1.2 Grade 4

Table 2 shows the changes in the degree of affection of the grade 4 students in all disciplines throughout the academic year. It can be seen that after the STEM education, over the three tests, students did not change their affection for various subjects significantly. The best-liked subjects were mainly the four disciplines of science, computer, art, and moral education. The subject with the lowest degree of the affection of students was always English. In the second semester, the degrees of affection of students for science, English, and sports were significantly lower than in the first semester. Compared with the pretest of the first semester, after the second semester, the degrees of affection for Chinese, computer, moral education, and art improved significantly ($p < 0.05$), but the degrees of affection for science and English decreased significantly, which may be related to the actual teaching situation.

4.1.3 Grade 6

Table 3 lists the changes in the degrees of affection of sixth-year students for different subjects. From the perspective of the whole school year, the degrees of

affection of grade 6 students for the subjects did not change significantly, indicating that the grade 6 students' understanding of various subjects was relatively stable. However, from the semester perspective, after the last semester, the attitudes of middle-school students to various subjects improved and the degrees of love for mathematics and Chinese subjects were more significant ($p < 0.05$). However, compared with the pretest of the first semester, in the posttest of the second semester, the degrees of affection for sports and other disciplines declined. Grade 6 students always face test pressures, which may also lead to a decline in performance in many subjects.

4.2 Learning attitudes toward STEM subjects

4.2.1 Grade 2

Emotional attitudes of grade 2 are examined from four dimensions: cooperative communication, solving problems, hands-on practice and career orientation. The changes in the mean value of emotional attitudes of second-year students are illustrated in Fig. 1. The data exhibited no significant differences in students' cooperative communication, solving problems, hands-on practice and career orientation. However, from the overall picture, an upward trend can be observed in these four dimensions. Because

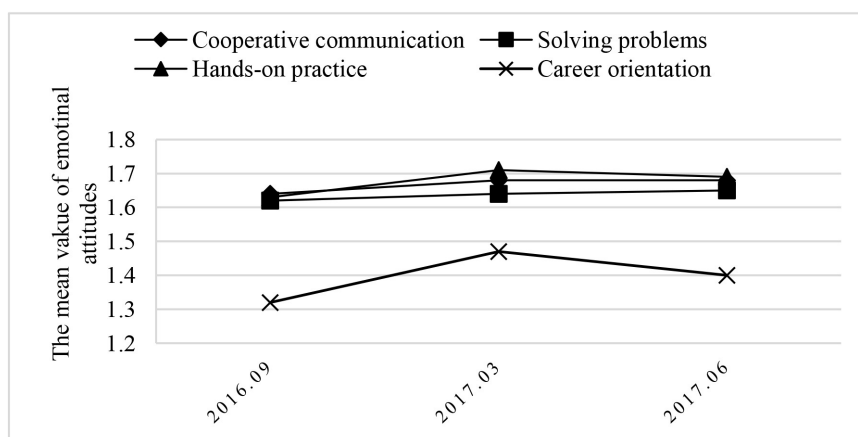


Fig. 1. Changes in the mean value of emotional attitudes of second-year students.

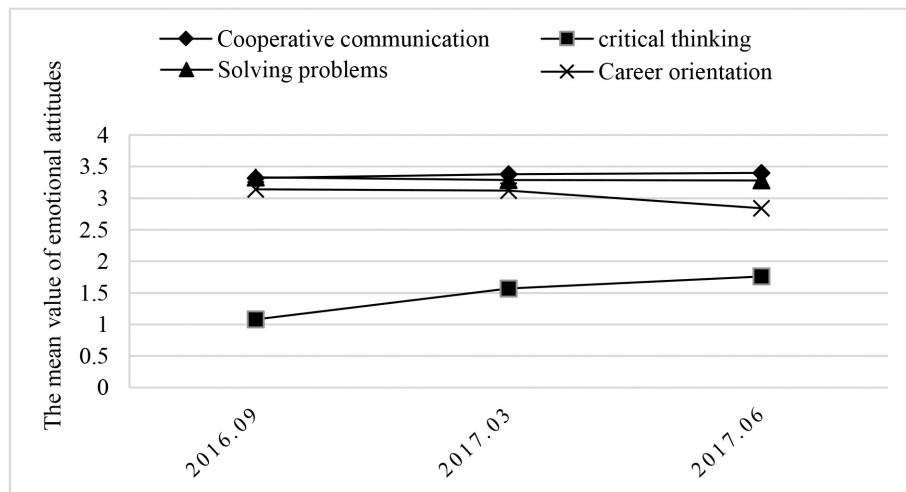


Fig. 2. Changes in the mean value of emotional attitudes of fourth-year students.

the career orientation average is relatively low, the analysis is that the course available for this semester has a certain role in promoting students' emotional attitudes.

4.2.2 Grade 4

The grade 4 students' emotional attitude scale evaluates students from four aspects: cooperative communication, critical thinking, solving problems, and career orientation. Students can choose one of the five levels of "very consistent," "comparative," "general," "not consistent," and "very non-conforming" (scored with 0 to 4 points).

It can be seen from Fig. 2 that the students' critical thinking scores continue to improve. The second semester after the test has significantly improved compared with the previous semester, indicating that the STEM course helps students to think critically. However, the overall score is still the lowest of the four and the future training space is

still very large. The scores of career orientation in the posttest are significantly lower than those in the previous semester. The data exhibited no significant change from the scores of the previous semester and the degree of affection of science in the postcourse semester also exhibits a significant decline. The analysis of this result is related to the theme of the semester.

4.2.3 Grade 6

The grade 6 emotional attitude scale evaluates students from four aspects: cooperative communication, critical thinking, solving problems and career orientation. Students can choose between five levels: "very consistent," "compliance," "general," "relatively inconsistent" and "very non-conforming" (corresponding to 0~4 points). The invalid questionnaires with all the same answers and blanks removed are analyzed as follows.

As can be seen from Fig. 3, over time, the grade 6

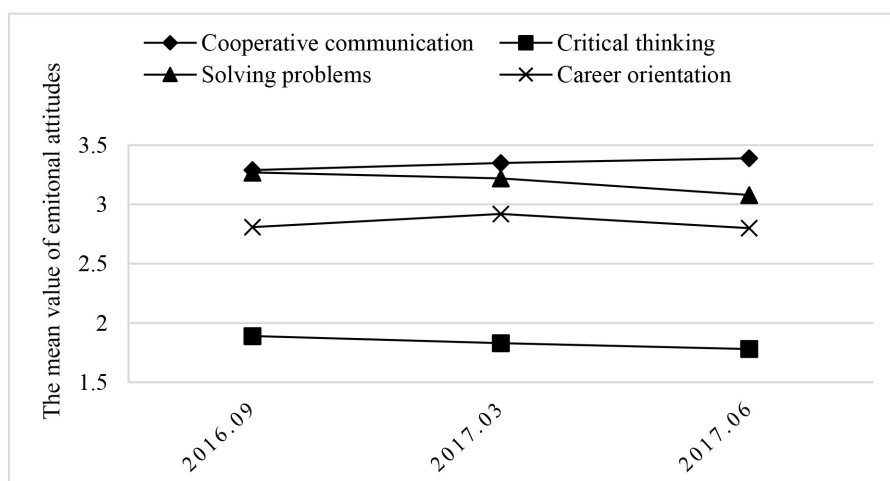


Fig. 3. Changes in the mean value of emotional attitudes of grade 6 students.

students do not differ significantly in the four dimensions of emotional attitudes. From start of the school year, the grade 6 students' cooperative communication attitude significantly improved and their critical thinking ability significantly declined. This shows that through the study of STEM projects, students' ability to communicate with each other is promoted, but the changes in critical thinking, solving problems and career orientation must be further understood, especially the reduction of critical thinking ability. This also explains from the side that career orientation may be relatively changeable in young students; if that interpretation is true, it explains that schools and parents must pay more attention to each other in teaching and encourage students to think and develop more.

4.3 *Disciplinary cognition of STEM*

4.3.1 *Grade 2*

The grade 2 curriculum theme mainly examines the students' understanding and application of relevant basic knowledge and joins the investigation of

engineering design thinking. The proportions of the four dimensions of science, technology, mathematics, and engineering in the last semester are 55%, 11%, 20% and 14%. The proportions of the four dimensions of science, technology, math, and engineering in the next semester were 40%, 30%, 20% and 10%. The changes in the scores of each knowledge point for grade 2 in the last semester are graphed in Fig. 4. The grade 2 students have significant improvement in the two dimensions of science and engineering in the first semester, whereas the dimension of math exhibits a significant decline and the dimension of technology reveals no significant changes. Overall, the students' cognitive scores exhibit significant improvement, which in turn explains that these courses are useful for teaching students to master the relevant knowledge points.

In addition, it can be seen from Fig. 5 that the grade 2 students have significantly improved the scores of engineering and science in the second semester, indicating that students have learned relevant knowledge applications through the

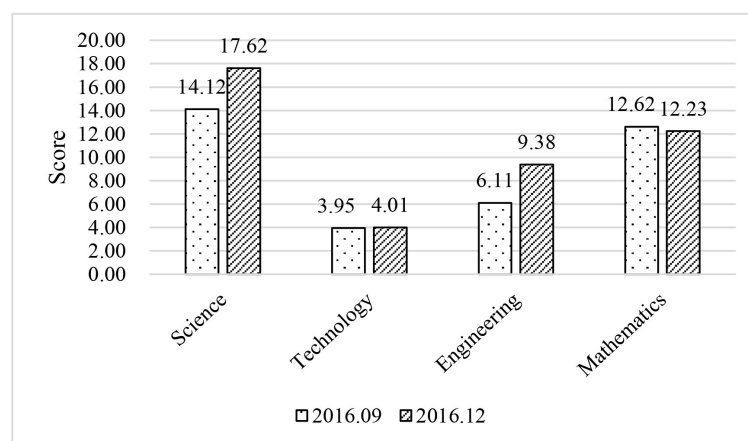


Fig. 4. Changes in the scores of each knowledge of grade 2 students in the first semester.

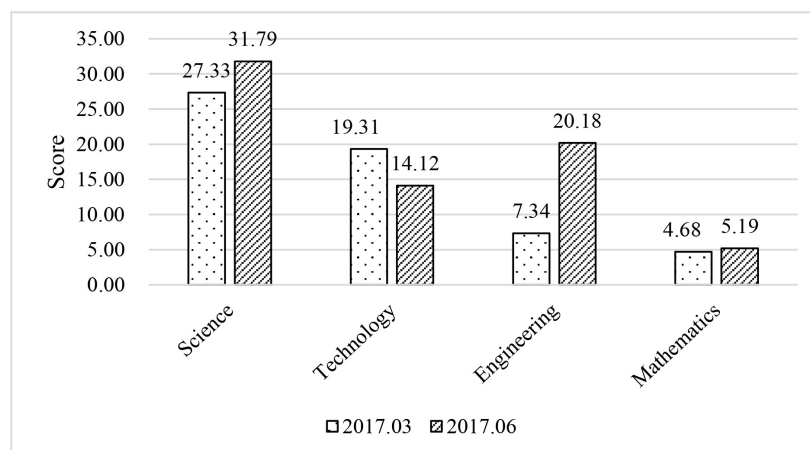


Fig. 5. Changes in the scores of each knowledge of grade 2 students in the next semester.

STEM courses. The scores of math do not increase significantly. The scores of the pretest and posttest are basically stable at very low scores, indicating that mathematics mastery was still hampered by grave challenges, and mathematics training was not sufficient. In addition, the data exhibited a significant downward trend in the score of the technology part. These problems were also related to the long time period of project scoring rate improvement.

4.3.2 Grade 4

The disciplinary cognition tests of grade 4 students center on the curriculums to examine students' understanding and application of technical knowledge of related disciplines and engineering design thinking ability. The proportions of the four dimensions of science, technology, mathematics and engineering in the last semester are 50%, 10%, 10% and 30%. The proportions of the four dimensions in the next semester are 50%, 20%, 10% and 20%. The changes in the scores of the grade 4 students at each knowledge point are depicted in Fig. 6 and Fig. 7.

It can be seen from Fig. 6 that the scores of students have been significantly improved in the science, technology, engineering and mathematics, especially in science and technology. Furthermore,

the performance of students in mathematical studies slightly improved; the school has considerable room for improvement in the training of students' mathematical knowledge. Moreover, this phenomenon has a certain relationship with the test questions of the semester.

In addition, it can be seen from Fig. 7 that in the second semester, students significantly improved their levels of science, technology, engineering and mathematics. Obviously, the improvements of students in these four dimensions are very large, which indicates the learning results of the relevant STEM courses.

4.3.3 Grade 6

The disciplinary cognition tests of grade 6 centered on the content of the textbooks, mainly examining the students' understanding and application of the basic mechanics of physics. Among them, the proportion of physics subject knowledge (science), engineering and math in the three dimensions of the previous semester are 40%, 20% and 10% respectively. In the four dimensions of the second semester, the proportions of science, technology, engineering and math are 50%, 20%, 20% and 10% respectively. The grade 6 students change the scores

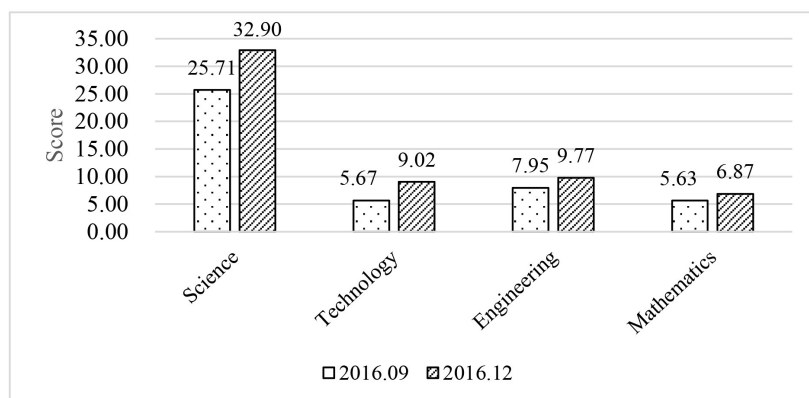


Fig. 6. Changes in the scores of each knowledge of grade 4 students in the first semester.

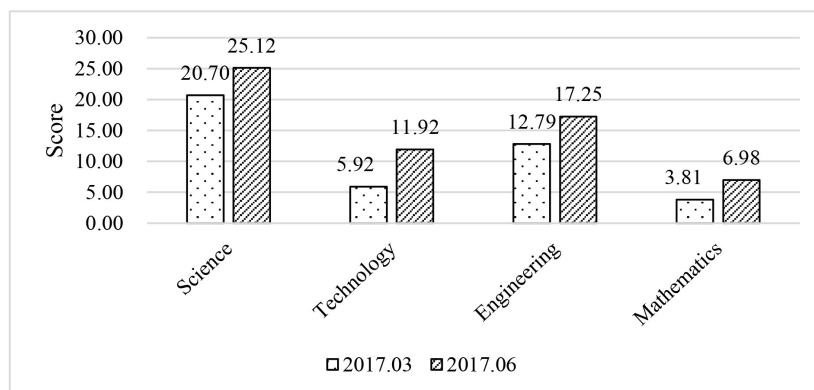


Fig. 7. Changes in the scores of each knowledge of grade 4 students in the second semester.

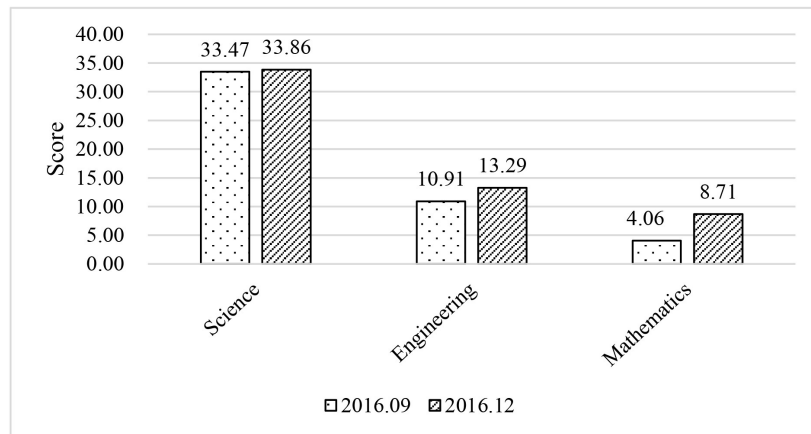


Fig. 8. Changes in the scores of each knowledge of grade 6 students in the first semester.

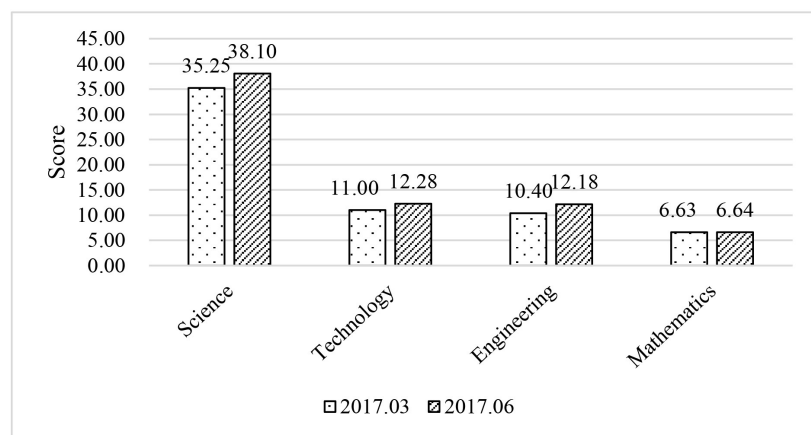


Fig. 9. Changes in the scores of each knowledge of grade 6 students in the second semester.

before and after the scores of each knowledge point as shown in Fig. 8 and Fig. 9. The students have significantly improved their engineering and math in the first semester. After the next semester, students have significantly improved their scientific knowledge and the course has achieved perfect results. The two semesters differ in the dimensions of improvement and the analysis has considerable relevance to the subject matter of the textbook.

4.4 Investigation of engineering design scenarios in the STEM disciplines

In the engineering design, one class was selected for each grade and scored respectively by three raters. The Kendall coefficient was used to test the consistency of the scores of the three raters and to ensure objective criteria for the following scores. The three levels of engineering design were divided into three evaluation dimensions, clarifying the problem, solving the problem and engineering design. The following is a comparative analysis of the overall situation of pretest and posttest in grade 2, grade 4 and grade 6 scenarios.

4.4.1 Grade 2

The engineering design scenario questions of grade 2 require students to design feasible solutions according to the scenarios. The situational questions mainly examine the students' engineering design ability from three dimensions and four levels. The three dimensions are clarifying problems, solving problems and engineering designing. The grades represent the "blank or wrong," "basic or partial completion," "effective problem solving" and "outstanding design ability" of the students' engineering design ability. The comparison of the engineering design scenario questions of the grade 2 students are shown in Table 4.

After the study of STEM project, the overall ability of second-year students in engineering design scenarios significantly improved. Students exhibited an upward trend in problem solving and engineering design, indicating that after STEM project learning, student's abilities to solve problems and engineering design significantly improved. However, the data exhibited a significant decrease in the dimension of the clarifying problems

Table 4. Comparison of the pretest and posttest of the grade 2

	2016.09(T1)		2016.12(T2)		2017.06(T3)		T2-T1	T3-T1
	Mean	Std	Mean	Std	Mean	Std		
Clarifying problems	0.44	0.68	0.55	0.93	1.11	0.91	-0.19*	-0.09
Solving problems	0.48	0.62	0.61	0.75	0.90	0.79	0.12	0.25*
Engineering designing	0.59	0.66	0.76	0.73	0.65	0.68	0.26*	0.17

Table 5. Comparison of the pretest and posttest of the grade 4

	2016.09(T1)		2016.12(T2)		2017.06(T3)		T2-T1	T3-T1
	Mean	Std	Mean	Std	Mean	Std		
Clarifying problems	0.23	0.03	0.47	0.04	0.97	0.74	0.21*	0.74
Solving problems	0.86	0.05	1.20	0.05	0.51	0.68	0.28	-0.36
Engineering designing	1.00	0.05	1.41	0.05	0.84	0.65	0.35	-0.15

Table 6. Comparison of the pretest and posttest of grade 6

	2016.09(T1)		2016.12(T2)		2017.06(T3)		T2-T1	T3-T1
	Mean	Std	Mean	Std	Mean	Std		
Clarifying problems	1.10	0.72	0.93	0.70	1.03	0.59	-0.19*	-0.09
Solving problems	0.85	0.83	0.97	0.91	1.10	0.85	0.12	0.25
Engineering designing	1.18	0.68	1.44	0.90	1.35	0.73	0.25*	0.17

($p < 0.05$), indicating that students lack exercise in this aspect. In the following STEM course, students should be trained in the ability to clarify problems.

4.4.2 Grade 4

The engineering design scenario of grade 4 requires students to design two practical devices, draw sketches, write design plans and engage with practical methods. This serves mainly to examine students' engineering design capabilities from three dimensions: clarifying problems, solving problems and engineering design. The engineering design scenario questions of the grade 4 students are compared in Table 5.

It can be seen from Table 5 that after the STEM course, the overall ability of students in engineering design slightly improved. In particular, students significantly improved in clarifying problems ($p < 0.05$). Students exhibited a rising and then downward trend in solving problems and engineering design dimensions. Therefore, after the STEM project learning, students' abilities in these two aspects were basically stable. The data did not exhibit noteworthy fluctuation, indicating that the first semester course was more effective than the second semester, and the second semester course theme can be considered for appropriate adjustment.

4.4.3 Grade 6

The engineering design scenario of grade 6 provides students with different outdoor scenes. Students are

required to design a solution to the problem according to the environmental conditions of the scene, draw the basic structure of the scheme model and explain the reasons for selection, scientific rationale and design ideas.

As can be seen from Table 6, the overall ability of grade 6 students in the engineering design situation questions significantly improved. In terms of clarifying questions, the two-semester middle-school students showed a downward trend; during the first semester a significant decline occurred after the test ($p < 0.05$). Given that the pretest and posttest topics are the same, due to lack of time, the students did not think deeply and elaborately on how to think about the steps to solve the problem. In the two semesters, students significantly improved their problem solving and engineering design dimensions, with students significantly improving their scores after the first semester ($p < 0.05$). However, some problems persisted: 1. Students' ideas were relatively simple and they were reluctant to think more deeply about actual and practical solutions. 2. Drawings lacked dimensional data and few students mentioned and used the word "quantity."

5. Conclusions

This study used data from a large sample ($N = 693$) of students to design a set of pretest and posttest evaluation questionnaires based on research questions and textbook content; the study developed a

set of STEM curriculum evaluation scales; these scales can evaluate STEM programs from multiple perspectives. The students of three grades liked STEM courses and claimed that they had learned considerable engineering and scientific knowledge. In addition, students' levels of emotional attitudes, subject cognition, and engineering design skill improved significantly, but substantial room for improvement remained. The advantages of this evaluation system can be summarized as follows:

- (1) The STEM course in the text contains numerous topics; in terms of these topics, the teaching effects of STEM projects can be evaluated comprehensively and systematically. Each semester course is guided by engineering design.
- (2) In the process of studying courses, students learn related subject knowledge, understand the engineer's occupation, and conduct experimental inquiry and other activities around specific topics.
- (3) This study investigated the questionnaire data from a large set of students, and compared the differences between the students in different grades before and after the STEM course. The analysis of the STEM curriculum has a certain promotion effect on the STEM ability of the students.

Relative to traditional single-factor survey forms, the evaluation method proposed in this paper can compare the feedback results of the pretest and posttest, and propose corresponding improvement suggestions based on the implementation effect of STEM education, which could train students' comprehensive quality ability. The disadvantage of the proposed method is that when large STEM courses are implemented on three levels, the teacher's STEM levels are uneven, which hindered the accuracy of the assessment.

Acknowledgments—This work was supported by the National Education Science “13th Five-Year plan” key research topics of Ministry of Education [grant number DCA170309].

References

1. T. Huziak-Clark, T. Sondergeld, M. Staaden, C. Knaggs and A. Bullerjahn, Assessing the Impact of a Research-Based STEM Program on STEM Majors' Attitudes and Beliefs, *School Science & Mathematics*, **115**(5), pp. 226–236, 2015.
2. M. Steiner, J. Kanai, C. Hsu and E. H. Ledet, Preparing engineering students for professional practice: using capstone to drive continuous improvement, *International Journal of Engineering Education*, **31**(1), pp. 154–164, 2015.
3. S. Han, R. Capraro and M. M. Capraro, How science, technology, engineering, and mathematics (STEM) project-based learning(PBL) affects high, middle, and low achievers differently: the impact of student factors on achievement, *International Journal of Science & Mathematics Education*, **13**(5), pp. 1089–1113, 2015.
4. C. J. Chung, C. Cartwright and M. Cole, Assessing the Impact of an Autonomous Robotics Competition for STEM Education, *Journal of STEM Education Innovations & Research*, **15**(2), pp. 24–34, 2014.
5. H. Burte, A. L. Gardony, A. Hutton and H. A. Taylor, Think3d!: Improving mathematics learning through embodied spatial training, *Cognitive Research*, **2**(1), pp. 13–30, 2017.
6. I. Cantley, M. Prendergast and F. Schlindwein, Collaborative cognitive-activation strategies as an emancipatory force in promoting girls' interest in and enjoyment of mathematics: A cross-national case study, *International Journal of Educational Research*, **81**, pp. 38–51, 2017.
7. K. Peterman, R. Kermish-Allen and G. Knezek, Measuring student career interest within the context of technology-enhanced stem projects: a cross-project comparison study based on the career interest questionnaire, *Journal of Science Education & Technology*, **25**(6), pp. 833–845, 2016.
8. C. Parker, E. L. Smith, D. McKinner and A. Laurier, The application of the engineering design process to curriculum revision: a collaborative approach to stem curriculum refinement in an urban district, *School Science and Mathematics*, **116**(7), pp. 399–406, 2016.
9. M. Gilliam, P. Jagoda, C. Fabiyi, P. Lyman, C. Wilson, B. Hill and A. Bouris, Alternate reality games as an informal learning tool for generating STEM engagement among underrepresented youth: a qualitative evaluation of the source, *Journal of Science Education & Technology*, **26**(3), pp. 295–308, 2017.
10. S. S. Guzey, M. Harwell, M. Moreno, Y. Peralta and T. J. Moore, The impact of design-based STEM integration curricula on student achievement in engineering, science, and mathematics, *Journal of Science Education and Technology*, **26**(2), pp. 207–222, 2017.
11. T. Tyler-Wood, G. Knezek and R. Christensen, Instruments for Assessing Interest in STEM Content and Careers, *Journal of Information Technology for Teacher Education*, **18**(2), pp. 345–368, 2010.
12. F. Castronovo, P. N. V. Meter, S. E. Zappe, R. M. Leicht and J. Messner, Developing problem-solving skills in construction education with the virtual construction simulator, *International Journal of Engineering Education*, **33**(2), pp. 831–846, 2017.
13. S. C. Fan and K. C. Yu, How an integrative STEM curriculum can benefit students in engineering design practices, *International Journal of Technology & Design Education*, **27**(1), pp. 107–129, 2015.
14. A. A. Wilson, E. Smith and D. L. Householder, Fundamental literacies required of adolescents' engineering designs: A multiple case study, *Annual meeting of the American Educational Research Association*, San Francisco, April, 2013.
15. M. Harwell, M. Moreno, A. Phillips, S. S. Guzey, T. J. Moore and G. H. Roehrig, A study of STEM assessments in engineering, science, and mathematics for elementary and middle school students, *School Science and Mathematics*, **115**(2), pp. 66–74, 2015.
16. S. Thompson and J. Lyons, Engineers in the Classroom: Their Influence on African-American Students' Perceptions of Engineering, *School Science & Mathematics*, **108**(5), pp. 197–211, 2010.
17. R. Ross, J. Whittington and P. Huynh, LaserTag for STEM Engagement and Education, *IEEE Access*, **5**, pp. 19305–19310, 2017.
18. R. M. Capraro, M. M. Capraro, J. J. Scheurich, M. Jones and J. Morgan, Impact of sustained professional development in stem on outcome measures in a diverse urban district, *The Journal of Educational Research*, **109**(2), pp. 181–196, 2016.
19. C. J. Chung, C. Cartwright and M. Cole, Assessing the Impact of an Autonomous Robotics Competition for STEM Education, *Journal of STEM Education Innovations & Research*, **15**(2), pp. 24–34, 2014.
20. T. J. Kennedy and M. R. L. Odell, Engaging Students in STEM Education, *Science Education International*, **25**(3), pp. 246–258, 2014.
21. L. J. Berk, S. L. Muret-Wagstaff, R. Goyal, J. A. Joyal and J. A. Gordon, Inspiring careers in STEM and healthcare fields through medical simulation embedded in high school science

- education, *Advances in Physiology Education*, **38**(3), pp. 210–215, 2014.
22. S. Jamieson, Likert scales: how to (ab)use them, *Medical education*, **38**(12), pp. 1217–1218, 2004.
 23. E. Saxton, R. Burns, S. Holveck, S. Kelley, D. Prince, N. Rigelman and E. A. Skinner, A Common Measurement System for K-12 STEM education: Adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking, *Studies in Educational Evaluation*, **40**, pp. 18–35, 2014.
 24. J. Bond, Y. Wang, C. Sankar, P. K. Raju and Q. Le, Female and minority students benefit from use of multimedia case studies, *International Journal of Engineering Education*, **30**(2), pp. 343–359, 2014.
 25. H. A. Alblooshi and L. May, Engaging women to study STEM through empowerment: A case from the United Arab Emirates (UAE), *2018 IEEE Aerospace Conference*, Big Sky, MT, USA, March 3–10, 2018.
 26. J. T. E. Richardson and E. King, Gender differences in the experience of higher education: Quantitative and qualitative approaches, *Educational Psychology*, **11**(3–4), pp. 363–382, 1991.
 27. B. Fralick, J. Kearn, S. Thompson and J. Lyons, How Middle Schoolers Draw Engineers and Scientists, *Journal of Science Education and Technology*, **18**(1), pp. 60–73, 2009.
 28. T. Stripling, Research projects funded. (Research, Education, and Practice Guidelines), *Educational Research Quantitative Qualitative & Mixed Approaches*, **5**(3), pp. 167–179, 2014.
 29. K. B. Wendell and C. Rogers, Engineering Design-Based Science, Science Content Performance, and Science Attitudes in Elementary School, *Journal of Engineering Education*, **102**(4), pp. 513–540, 2013.
 30. S. A. Rodzalan and M. M. Saat, The Perception of Critical Thinking and Problem Solving Skill among Malaysian Undergraduate Students, *Procedia—Social and Behavioral Sciences*, **172**(3), pp. 725–732, 2015.
 31. E. A. Plant, A. L. Baylor, C. E. Doerr and R. B. Rosenber-Kima, Changing middle-school students' attitudes and performance regarding engineering with computer-based social models, *Computers & Education*, **53**(2), pp. 209–215, 2009.
 32. G. Ragusa and C. T. Lee, The impact of focused degree projects in chemical engineering education on students' research performance, retention, and efficacy, *Education for Chemical Engineers*, **7**(3), pp. 69–77, 2012.
 33. K. H. Kim, Can we trust creativity tests? A review of the Torrance tests of creative thinking (TTCT), *Creativity Research Journal*, **18**(1), pp. 3–14, 2006.
 34. D. Yu, Y. Zhang, Y. Xu and C. Wang, Improvement in critical thinking dispositions of undergraduate nursing students through problem-based learning: a crossover-experimental study, *Journal of Nursing Education*, **52**(10), pp. 574–581, 2013.
 35. J. Qian and J. Yu, Students Creativity Modeling with Gene Expression Programming, *2012 International Conference on Computer Science and Electronics Engineering*, Hangzhou, China, March 23–25, 2012.

Qin Ni is currently an assistant professor in College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai, China. She received her PhD degree from Universidad Politécnica de Madrid in 2016. Her research interests include data mining, smart environment, activity recognition, assistive robot design and STEM education.

Lele Zhang is now a master student in College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai. Her research interest includes data analysis, social network community discovering.

Bo Zhang received the PhD degree in computer science from Tongji University in 2009. He is now professor of College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai, China. His current research interests include trust computation, and social network analysis. He is now a director of trust computation project, which is funded by National Nature Science Foundation of China.

Feng-kuang Chiang (corresponding author) is a distinguished professor in the Department of Educational Technology at Shanghai Normal University, China. His research focused on STEM education, learning space and integration of information technology and curriculum.