*Research Article in Assessment Method*

# Improving students' academic achievement and attitude in science through comprehensive evaluation

Zhenqiang Wang[1,2*], Xiaoyu Ni[1], and Wenbin Xu[1]

[1] Institute of Curriculum and Teaching, Nanjing Normal University, Nanjing, Jiangsu, China

[2] Nanjing Xiaozhuang College affiliated Primary School, Nanjing, Jiangsu, China

[*] Corresponding Author: 63112981@163.com

(Received: 04/26/2023; Accepted: 12/12/2023; Published: 12/18/2023)

DOI: https://doi.org/10.37906/real.2023.2

**Abstract**: Evaluation is to promote, evaluation is to develop. Based on the Student Evaluation Guidelines issued by the Ministry of Education of China as a theoretical framework, this study incorporated classroom performance, extracurricular performance, academic achievement and extended evaluation into the comprehensive evaluation of primary school science learning, and conducted eight years of practical exploration. The data of scientific attitude (classroom performance, extracurricular performance, extended evaluation) and academic performance of 178 students in Grade 3 for a period of 3 years were analyzed. Through multiple comparisons of descriptive statistics, analysis of variance and analysis of variance with Spss26, the research found that: (a) The use of comprehensive evaluation to improve pupils' academic performance in science presents a gradual increasing trend, and there is a significant difference during the school period. (b) Using comprehensive evaluation to improve pupils' attitude towards science learning, showing an overall growth trend. (c) The use of comprehensive evaluation to improve students' classroom performance and expansive evaluation plays a promoting role, showing a gradually increasing trend, and there is a significant difference during the study period. (d) Comprehensive evaluation has a certain impact on the improvement of students' extracurricular performance.

**Keywords**: Academic achievement in science, Comprehensive evaluation, Classroom performance, Extracurricular performance, Evaluation of extensibility, Scientific learning attitude

## 1. Introduction

With the development of science and technology, more and more international organizations start to pay attention to science education and emphasize the importance of scientific literacy. For example, the US Next Generation Science Standards (NGSS), the Program for International Student Assessment (PISA), Trends in International Mathematics and Science Research (TIMSS), and the OECD launched a new round of PISA tests in 2022 to add creative thinking (Grab au et al., 2022). PISA results reveal the differences in scientific literacy among students from different countries, which are mainly studied by means of paper-and-pencil tests (OECD, 2016; OECD, 2019). Science education standards around the world emphasize the importance of engaging students in science learning (NGSS, 2013). Evaluation is for promotion and evaluation is for development. In addition, in the process of comprehensive evaluation, it is necessary to consider the specific index of evaluation index and the weight of evaluation (Hwang et al., 2018; Yang et al., 2021).

In China, the Outline of the Action Plan for Scientific Literacy of the Whole People (2021-2035) points out that improving the scientific literacy of the whole people is of great significance to enhancing the country's independent innovation capacity and cultural soft power. In order to enhance the scientific quality of teenagers, the level of science education in basic education should be improved (China, 2021). However, primary school science is a practical course, and evaluation only through paper-and-pencil tests cannot truly detect students' scientific literacy. It is necessary to strengthen students' process learning and encourage them to add practical evaluation. Therefore, comprehensive evaluation should be carried out in many aspects to evaluate students more comprehensively. The compulsory education science curriculum standards advocate the diversified learning methods based on inquiry and practice, so that students can actively participate in the process of scientific inquiry and technology and engineering practice with their hands and minds (China, 2022).

At present, the biggest obstacle to the implementation of integrated science curriculum is the problem of teachers, which has become a bottleneck restricting the implementation of integrated science curriculum in primary and secondary schools and even a new round of curriculum reform. The main reason is that the role of comprehensive science teachers is disorganized (Xu, 2007). Traditional evaluation of students' learning usually takes their scores of unit tests, mid-term exams and closed-book written tests of final exams as the fundamental reference. This single and terminal evaluation not only shows various defects in students' comprehensive ability and individual strengths, but also fails to diagnose students' operational ability, oral expression ability, language communication ability, independent cooperative exploration ability, and even fails to accurately evaluate students' learning process and development process (Lin, 2013). Science curriculum focuses on cultivating students' learning interest and developing students' various abilities. The development of science curriculum needs a matching curriculum and teaching evaluation system. However, at present, the science curriculum has not established a whole set of evaluation system, which makes the teachers who are used to exam-oriented education are full of doubts about the teaching of science curriculum. This undoubtedly affects the reform process and implementation effect of science curriculum (Pan, 2011). Comprehensive quality assessment is an observation, record and analysis of students' overall development, an important means to discover and cultivate students' good personality, and an important system to further promote quality education (Wang, 2020).

This study aims to better implement the original intention of curriculum standard adjustment, improve students' academic achievement and attitude in science, and use comprehensive evaluation to improve pupils' academic achievement and attitude in science research.

## 2. Literature review

### 2.1. Comprehensive evaluation

Comprehensive evaluation is the observation, record and analysis of students' overall development, which is an important means to find and cultivate students' good personality. Grade evaluation is not only to evaluate students' tests or assignments, but also to establish a specific standard and scale (Barbara et al., 1998; Wang, 2020). The General Plan issued by the Ministry of Education of China emphasizes the reform of "four evaluations", including "process evaluation", "value-added evaluation", "comprehensive evaluation" and "peer expert review mechanism" (China, 2020). There have been some theoretical studies and discussions on the establishment of the index system of comprehensive quality evaluation and the selection of comprehensive evaluation methods, but they only involve college students (Huang, 2001; Ke, 2010; Wu et al., 2000). Wang et al. (2019) conducted a study based on the correlation factors between academic achievement and comprehensive quality evaluation.

There are few researches on the practicability of comprehensive evaluation of students, especially the empirical researches, especially the results of comprehensive evaluation of specific disciplines. Based on this, this study aims to provide inspiration and reference for improving pupils' academic achievement and attitude in science.

2.2. Academic achievement in science

PISA is held every three years. It mainly monitors the academic development of different countries, reflects the learning effect of students by comparing the academic development of different countries, and promotes the development of students' knowledge and skills based on the learning effect (Chen et al., 2015). PISA mainly tests students' mastery of knowledge and skills, and focuses on whether students can use the knowledge and skills they have acquired to meet various challenges in future life (Chen et al., 2015). Each time TIMSS conducts evaluation, it will have its evaluation focus. Take TIMSS 2015 targeted Grade 4 and 8 students as an example, it focuses on science content, cognition and practice. (Mullis et al., 2018), observing five main aspects of raising questions, forming evidence, processing data, answering research questions, and scientific practice (Jones et al., 2015). In the United States, NEAP is held once a year and started in 1969 (Lee et al., 2010). It mainly evaluates the academic development of students in grades four, eight and twelve, and its evaluation content is more than that of PISA and TIMSS(Randy,2018). Taking NAEP 2015 as an example, the evaluation mainly consists of two parts: scientific content and scientific practice (NCES, 2019).

Since 2016 in Zhejiang Province of China in 2016, science academic examination results have been graded as A, B, C, D that represent pass, and E that means failure. Grades A, B, C are delimited according to the accumulative proportion of 15%, 30%, 30% (Jin, 2020). Flexible and diverse forms of examination evaluation are conducted in Japan. For example, small essay writing, interviews, simulated teaching, seminars and so on are all widely used forms of enrollment (Weng, 2017). Shanghai, China introduced "green evaluation", which mainly introduced learning motivation, academic burden, teacher-student relationship, moral behavior, physical and mental health and other indicators (Li, 2016). China only proposed process and formative evaluation in the 2017 primary school science curriculum evaluation proposal, and added primary school science academic proficiency test to the newly published curriculum standards in 2022 (China, 2017; China, 2022). Primary school science scholastic proficiency test is a new thing in China, which needs to learn from the United States and other countries in both the content and form of the test.

2.3. Scientific learning attitude

Gardner (1975) divided scientific attitudes into "scientific attitudes towards " and "attitudes towards science". The "scientific attitude" is not a single structure, but contains many different dimensions. Some scholars divided the attitude towards science into seven dimensions: in-school science learning, out-of-school science learning, science practice, future participation in science, importance of science, general interest in science and self-concept of science (Ajzen,2001; Kind et al., 2007). Other scholars define a scientific attitude as an interest in scientific topics, a fondness for science, and the value of science (Chi & Wang, 2017).

In recent years, PISA has adopted the Likert Scale to evaluate students' attitude towards science. In 2015, PISA test and TIMSS divided scientific attitude into three dimensions: interest in science, value of evaluating scientific inquiry methods and environmental awareness. (OECD, 2017; Wein burgh et al., 2000)

Tee et al.(2018) found that there are significant differences between boys and girls in their attitudes towards science, and boys have a more positive attitude towards science than girls. Guo et al. (2022) not only found that there is no significant difference of practical significance in students' attitudes towards

science between different genders; There are significant differences in the attitudes of students with different academic levels towards science, and students with higher academic levels have more positive attitudes towards science.

**3. Evaluation of process**

William, a British scholar, found through long-term classroom observation and investigation that it can stimulate learning motivation, provide feedback to promote learning, and encourage learners to become the master of their own learning (Dylan Wiliam, 2019). Koji Tanaka, a professor at Kyoto University in Japan, carried out the process evaluation practice through the evaluation methods of "study topic method" and "one-page portfolio evaluation", and achieved good teaching effects (Tian & Xiang, 2012). Yang (2015) integrated hierarchical teaching strategies with procedural evaluation, and found through practice that most students' enthusiasm and self-confidence were significantly improved, and most students' ability to find problems and solve problems collaboratively was also strengthened.

Astin A W (1999) found that students' active participation in classroom learning is positively correlated with learning gains. Larraz (2017) believes that cooperative learning can improve students' self-discipline, and they develop and improve in leadership, teamwork, reflection and other aspects. In view of the phenomenon of low achievement in social science, some scholars use jigsaw learning method to carry out cooperative teaching to promote the skills of teachers and students (Yalvema et al., 2015; Ching, 2015). organized students into study groups in science class, and found that students' scientific learning ability and emotional experience had been improved and accumulated (Zhao, 2017).

Habeshaw (1986) found that students could complete their homework well by using diaries to display their homework. Astin (1996) found that cooperation among students has an incentive effect on teaching. Perry, Menec & Struthers (1996) found that cooperative learning can increase students' awareness and self-control in the classroom. Chickering & Gamson, (1987) Cooperative learning can also reduce students' sense of self-control and increase their anxiety.

3.1. Extracurricular performance

Emsen & Perihan (2011) found that students' attitudes towards research assignments were significantly different depending on gender, class level and school type. Middle school students showed no meaningful differences in their attitudes toward gender-specific research assignments. Hasche et al. (2000) found no difference in student achievement in comprehensive or traditional homework. Compared with students with traditional homework, students with integrated homework had significantly lower stress levels at home, and they had significantly more positive attitudes toward school. However, their individual need for help with homework was higher.

Grolnick et al. (1994) believed that all parents' investment in education, including time and energy, is for the learning and development of their children. Patallea et al. (2008) found that parents' participation in students' homework behavior can ultimately promote students' homework completion, which can significantly improve students' academic performance in primary school. Of course, there is the opposite situation. Xu et al. (2010) conducted an observation study on parent-child reading of fifth-grade students and found that when parents participated in students' reading, students' scores declined. Pezdek (2002) found that the time parents spent on their children's math homework had no relationship with their children's math performance. In the past few decades, there have been many researches on the effect of parents' involvement in homework on students' academic performance, but the results have not been consistent.

### 3.2. Evaluation of extensibility

In 1824, German chemist Liebig set up a special teaching laboratory in Giessen University, allowing all students to do experiments themselves, which was the pioneer of experimental teaching method (Ruegg, 2004). China's Science Curriculum Standards for Primary Schools of Compulsory Education clearly put forward educational objectives such as experimental design, obtaining information through experiments and being willing to participate in experiments. Among the 18 learning contents, it is suggested that 14 contents should be carried out in the form of experimental activities, and it is required that the experiments should be included in formative evaluation and terminal evaluation. As for education places, both laboratories and classrooms are two "main learning places" (China, 2017). Opinions on Strengthening and Improving Experimental Teaching in Primary and Secondary Schools issued by the Ministry of Education of China in November 2019 once again emphasized the need to consolidate the foundation and fully carry out experiments stipulated by national curriculum standards (China, 2021). Wang et al. (2021) mainly analyzed that science experiment teaching would not only have a significant impact on students' science achievement, but also promote students' understanding of scientific knowledge, scientific inquiry ability and interest in science learning. Zhang & Tang (2017) Students' self-selection and participation in school-organized extracurricular activities have positive or indirect effects on students' science achievement.

In this study, the comprehensive evaluation is divided into process evaluation accounting for 40%, academic performance evaluation accounting for 60%, and extended evaluation as additional results, which mainly includes students' participation in science and technology associations, science and technology festivals, vacation science experiments, and provincial and municipal science and technology competitions. The process evaluation includes classroom performance accounting for 20%, and extracurricular performance accounting for 20%. Classroom performance mainly includes students' personal speech in class, teamwork in group performance, and experimental operation. Extracurricular performance mainly includes students' personal homework and daily science experiment homework. Use comprehensive evaluation to improve pupils' academic achievement and attitude in science, improve students' classroom performance, extracurricular performance and extended evaluation.

### 3.3. Research Focus

This study focuses on the practical exploration of comprehensive evaluation of science curriculum by 178 students from 4 classes in Grade 3 of a primary school in Nanjing, Jiangsu Province, China. The reason for this case is that there was no academic proficiency test in China's primary school science curriculum until 2022. The traditional evaluation of students' learning usually takes the scores of students' unit tests, mid-term examinations and closed book written examinations as the fundamental reference. Students' hands-on ability, oral expression ability, language communication ability and independent cooperative exploration ability cannot be clearly diagnosed (Lin, 2013). Science curriculum focuses on cultivating students' learning interest and developing students' various abilities. At present, a complete set of evaluation system has not been established in science curriculum (Pan, 2011).

Students usually do not pay attention to the attitude of science learning. In order to guide students to attach importance to the attitude learning of science curriculum, process evaluation and comprehensive evaluation will affect students' academic achievement and attitude of science. Therefore, in combination with China's Educational Evaluation Guidelines and the Evaluation Suggestions of Compulsory Education Curriculum Standards (China, 2022), this study constructs the comprehensive evaluation of primary school science learning, and uses the comprehensive evaluation to try to improve pupils' academic performance and attitude towards science. The attitude here mainly includes students' classroom performance, extracurricular performance and extended evaluation.

### 3.4. Research Aim and Research Questions

Using comprehensive evaluation to improve pupils' academic achievement and attitude in science. Comprehensive evaluation to improve students' classroom performance, extracurricular performance and evaluation of extended activities. The comprehensive evaluation in this study consists of three parts: process evaluation, academic achievement evaluation and extended evaluation. The process evaluation includes: Class performance account for 20% (individual performance and group performance), extracurricular performance account for 20% (written work, daily scientific experiments), homework account for 20% of the overall evaluation, group cooperation account for 20%, academic performance evaluation account for 60% (mainly to examine scientific knowledge, experimental design, inquiry practice, attitude responsibility, etc.), The extended evaluation mainly includes participating in school science and technology societies, science and technology festivals, science experiments during holidays, provincial and municipal science and technology competitions, etc. The final total score is based on the five levels of A, B, C, D and E. Therefore, this study focuses on using comprehensive evaluation to improve pupils' academic achievement and attitude in science. The following research questions are presented:

Using comprehensive evaluation to improve pupils' academic achievement and attitude in science.

1. Use comprehensive evaluation to improve students' academic performance.
2. Use comprehensive evaluation to improve students' attitude towards science learning.
3. Use comprehensive evaluation to improve students' classroom performance.
4. Use comprehensive evaluation to improve students' extracurricular performance.
5. Use comprehensive evaluation to improve students' extended evaluation.

### 3.5. Research Methodology

#### 3.5.1. General Background

Comprehensive evaluation is used to improve pupils' academic achievement and attitude in science. At present, the research of the comprehensive quality evaluation of Chinese students mainly focuses on the relevant theoretical level, the mode, the implementation criteria and so on. Mohammadpour (2013) found that students with a more positive attitude towards science generally achieve better academic performance in science. Wael et al. (2021) found that PBL project-style school-based cooperation can improve students' attitude towards science learning in school. Based on the Student evaluation guidelines and science curriculum standards issued by the Ministry of Education of China, this study constructed a comprehensive evaluation of primary school science learning with the theoretical framework, including classroom performance, extracurricular performance, academic performance and extended evaluation into the comprehensive evaluation index, and conducted an 8-year primary school science comprehensive evaluation practice exploration during 2014-2022 in a school in Nanjing, China. Focus on the third-grade students from the spring of 2020 to the end of the spring of 2022, a total of five semesters, the same group of 178 students in the attitude to science (classroom performance, extracurricular performance, extended evaluation), academic results of the statistical analysis. The results of the second semester of the third grade were taken as independent variables, and the academic performance of the second semester was taken as independent variables.

#### 3.5.2. Participants

A total of 178 Grade-3 students (98 boys and 80 girls) from 4 classes participated in the study. All students were taught by the same science teacher for three years, using the same teaching methods, teaching organization, evaluation methods and evaluation criteria. This science teacher has the title of

national senior science and technology counselor and a master's degree. Before the experiment, the teacher discussed with the teaching and research team and submitted the comprehensive evaluation plan to the Office of Teaching Affairs and the Office of the president of the school for approval. In the first week of school, all students and parents will be informed of the comprehensive evaluation method and the proportion. In addition, the research was also supported by the key project of Jiangsu Province's 13th Five-Year Plan and the key project of the Life and Practice Education Committee of China Tao Xingzhi Research Society.

3.5.3. Instrument and Procedures

The study was conducted in the fall of 2019 in four classes of Grade-3 at a primary school in Nanjing, China. Comprehensive evaluation from the classroom performance, extracurricular performance, academic performance, extended evaluation of several aspects. Class performance account for 20% (individual performance and group performance), extracurricular performance account for 20% (written work, daily scientific experiments), academic performance evaluation account for 60% (mainly to examine scientific knowledge, experimental design, inquiry practice, attitude responsibility, etc.), The extended evaluation mainly includes participating in school science and technology societies, science and technology festivals, science experiments during holidays, provincial and municipal science and technology competitions, etc. The final total score is based on the five levels of A, B, C, D and E. First of all, the four classes are taught by the same science teacher, and the teaching methods, teaching places, homework requirements, academic level content and participation opportunities are fair to every student. During this period, the same comprehensive evaluation method was adopted in each semester. At last, horizontal and vertical comparisons were made for the performance indicators of 178 students from the four classes in five semesters in three years.

3.5.4. Comprehensive evaluation of primary school science learning

From the fall of 2014, the researchers explored and practiced the comprehensive evaluation system based on the student evaluation guidelines and science curriculum standards issued by the Ministry of Education of China. After five years of modification and improvement of the evaluation system, the five-level comprehensive evaluation system was finally developed. The evaluation of students' learning quality should not only refer to their academic development level, but also take into account their specialty development and academic burden, so as to achieve both qualitative and quantitative evaluation (Liu, 2018). Comprehensive evaluation can be realized in various forms, such as independent enrollment, multi-factor quantitative synthesis of total scores, expansion of the comprehensive quality evaluation role of middle school within the unified entrance examination and enrollment system, and the combination mode of multi-factor evaluation determined independently by schools (Bian,2017). The General Plan issued by the Ministry of Education of China emphasizes the reform of "four evaluations", focusing on the construction of a world-class education evaluation system with Chinese characteristics, but "process evaluation", "value-added evaluation", "comprehensive evaluation" and "peer expert review mechanism" (China, 2021).

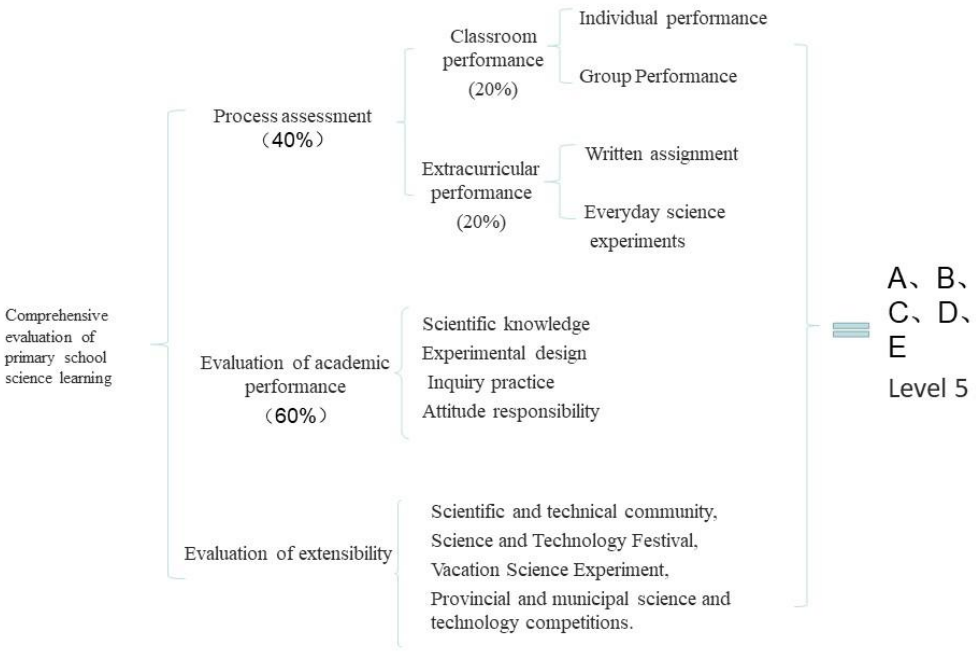


**Figure 1**. Comprehensive evaluation of primary school science learning

In this study, the comprehensive evaluation of primary school science learning is divided into three parts: process evaluation, academic achievement evaluation and extended evaluation. Among them, the process evaluation accounts for 40%, and the academic performance evaluation accounts for 60%. The extended evaluation is included in the comprehensive evaluation as an additional score, and the full score of the extended evaluation is 10. The extended evaluation is not included in the comprehensive evaluation as an incentive rating method that everyone must participate in. Finally, the cumulative results of these three parts are the comprehensive evaluation results, A⩾85,85>B⩾75,75>C⩾65,65>D⩾60,60>E.

Classroom performance mainly refers to students' performance in science class. It mainly includes individual performance and group performance, which are specifically reflected in individual speech in science class, design of classroom experiments conducted by groups, and scores of classroom science experiments. A semester consists of 18 weeks, with 2 science classes given week. The extra score of each class is no more than 4 points. After the semester, the score will be assigned according to the ranking.

Extracurricular performance mainly refers to the student's completion of work outside the science classroom. Here it mainly refers to the individual homework completed by students and some daily extra-curricular science experiments, which mainly reflects the homework related to the study of science courses completed by individuals. There are 20 assignments per semester, half of which are written and half of which are daily science experiments. For example, in the fifth grade, the written homework includes the classification of light sources and the experimental report of heat transfer in metals, etc., while the daily scientific experiments mainly include making a periscope, making diorama models of different terrains, making solar system models, ecological bottles, etc.

The evaluation of academic performance, accounting for 60%, mainly focuses on the core qualities in the curriculum standards, namely scientific knowledge, experimental design, inquiry practice, attitude responsibility, etc. (China, 2022) Paper and pencil test is adopted to focus on evaluating students' scientific inquiry ability, technology and engineering practice ability, and innovative ability to solve practical problems. Design questions mainly include fill-in-the-blank, choice, judgment, experimental design,

question and answer. The exam lasts 40 minutes and is closed book. The total score is 100 points. For example, see Annex 2 for end-of-term papers of a certain grade.

The expansive evaluation is included in the comprehensive evaluation as an additional score. The full score of the expansive evaluation is 10. The expansive evaluation is not included in the comprehensive evaluation as an incentive rating method, as everyone must participate in the completion. The extended evaluation in this study is mainly to give extra points to students who participate in science competitions at all levels. One point is added for winning school-level certificates, two points is added for winning district-level certificates, and three points is added for winning city-level certificates and above. Only the highest score is certified for the same competition. For example, students participate in the science and technology festival activities organized by the school: paper airplanes, gizmos, small papers, etc., model competition held in the city, Jiangsu Province science and Technology Innovation Competition, etc. See Annex 3 for some awards.

## 4. Data Analysis

Spss26 software was used to analyze and process the performance data of 178 students in class, extracurricular performance, academic performance and extended evaluation for five semesters in three years. The basic descriptive statistics of the values were determined, the mean difference test was conducted, and the five semesters were analyzed in terms of academic performance, classroom performance, extracurricular performance, and extended evaluation, and the significance was tested, and the post-test analysis was conducted.

### 4.1. Changes in students' academic performance tracked for three years

Descriptive statistical analysis

**Table 1**. Students' academic performance statistical description table

| | N | Least | Max | Mean | Variance | Skewness | | Kurtosis | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | S | E | S | E |
| 5-2 | 178 | 44 | 100 | 89.14 | 73.568 | -1.841 | .182 | 6.721 | .362 |
| 5-1 | 178 | 50 | 100 | 83.43 | 79.840 | -.704 | .182 | 1.662 | .362 |
| 4-2 | 178 | 36 | 96 | 81.20 | 102.897 | -1.277 | .182 | 2.434 | .362 |
| 4-1 | 178 | 42 | 95 | 77.44 | 99.378 | -1.060 | .182 | 1.790 | .362 |
| 3-2 | 178 | 34 | 91 | 75.12 | 86.907 | -1.162 | .182 | 2.503 | .362 |

*Note:* 5-2: The second term of Grade Five,5-1: The first term of Grade Five,4-2: The second term of Grade Fourth,4-1: The first term of Grade Fourth,3-2: The second term of Grade Three, S: Statistics, E: Error of standard

According to the data in Table 1, the average academic achievement of students in the second semester of Grade three is 75.12, the average academic achievement in the first semester of grade four is 77.44, the average academic achievement in the second semester of grade four is 81.20, the average academic achievement in the first semester of grade five is 83.43, and the average academic achievement in the second semester of grade five is 89.14. In general, the experimental class students' academic performance in science in five semesters showed a rising trend.

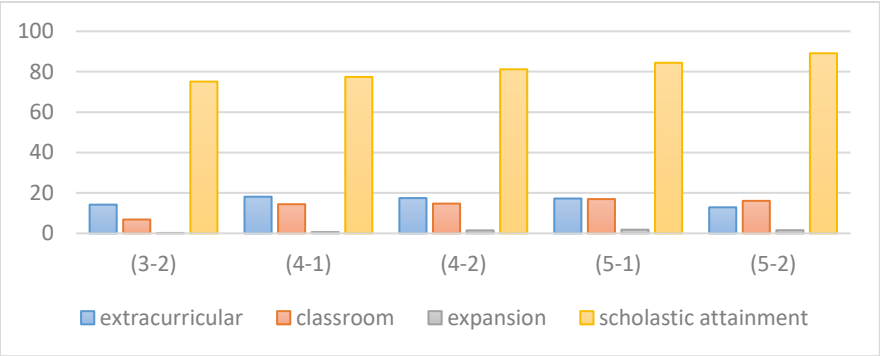


**Figure 2**: Comparison of the mean results of Students'

*Note:* 5-2: The second term of Grade Five,5-1: The first term of Grade Five,4-2: The second term of Grade Fourth,4-1: The first term of Grade Fourth,3-2: The second term of Grade Three

Variance analysis of academic achievement in science

**Table 2**. Test table for homogeneity of variance of academic achievement in science

| | | LS | 1 DOF | 2 DOF | p |
|---|---|---|---|---|---|
| FSFS | Based on average | 1.268 | 4 | 885 | .281 |
| | Based on the median | 1.117 | 4 | 885 | .347 |
| | Based on median and with adjusted degrees of freedom | 1.117 | 4 | 870.691 | .347 |
| | Based on the post-clipping average | 1.154 | 4 | 885 | .330 |

*Note: LS:* Levin's statistics, S: Significance, FSFS: Five semesters full sample

The variance analysis of students' academic performance in science for five consecutive semesters is carried out. First, the homogeneity test of variance is carried out, as shown in Table 2. Based on the mean Levin statistics, it is 1.268, significance P=0.281>0.05, homogeneity of variance.

**Table 3**. Results table of ANOVA analysis of variance for academic Achievement in Science

| | QS | DOF | MS | F | p |
|---|---|---|---|---|---|
| Interblock | 21209.647 | 4 | 5302.412 | 59.902 | .000 |
| Intra-class | 78338.242 | 885 | 88.518 | | |
| Total | 99547.889 | 889 | | | |

*Note:* QS: Quadratic sum, MS: Mean square, S:Significance

According to the results in Table 3, F=59.902, P<0.001. There are significant differences among the five groups of data, indicating that the students' academic performance in science in the five semesters has significant differences, which can be further compared after the event.

**Table 4.** Post-test of academic achievement in science

| Dependent Variable: academic achievement | | | | | | |
|---|---|---|---|---|---|---|
| **HSD** | | | | | | |
| | | | | | 95% confidence interval | |
| (I) VAR00002 | (J) VAR00002 | MD(I-J) | SE | P | LL | UL |
| 1 | 2 | -2.320 | .997 | .137 | -5.046 | .406 |
| | 3 | -6.084* | .997 | .000 | -8.810 | -3.358 |
| | 4 | -8.315* | .997 | .000 | -11.041 | -5.589 |
| | 5 | -14.022* | .997 | .000 | -16.749 | -11.296 |
| 2 | 1 | 2.320 | .997 | .137 | -.4058 | 5.046 |
| | 3 | -3.764* | .997 | .002 | -6.490 | -1.038 |
| | 4 | -5.994* | .997 | .000 | -8.720 | -3.268 |
| | 5 | -11.702* | .997 | .000 | -14.428 | -8.976 |
| 3 | 1 | 6.084* | .997 | .000 | 3.358 | 8.810 |
| | 2 | 3.764* | .997 | .002 | 1.038 | 6.490 |
| | 4 | -2.230 | .997 | .167 | -4.956 | .496 |
| | 5 | -7.938* | .997 | .000 | -10.664 | -5.212 |
| 4 | 1 | 8.315* | .997 | .000 | 5.589 | 11.041 |
| | 2 | 5.994* | .997 | .000 | 3.268 | 8.720 |
| | 3 | 2.230 | .997 | .167 | -.496 | 4.956 |
| | 5 | -5.708* | .997 | .000 | -8.433 | -2.982 |
| 5 | 1 | 14.022* | .997 | .000 | 11.297 | 16.749 |
| | 2 | 11.702* | .997 | .000 | 8.976 | 14.428 |
| | 3 | 7.938* | .997 | .000 | 5.212 | 10.6642 |
| | 4 | 5.708* | .997 | .000 | 2.981 | 8.4339 |

*. The significance level of the mean difference is 0.05.

*Note:* MD: Mean difference, SE: Standard error, LL: lower limit, UL: upper limit

As can be seen from Table 4, the academic performance is divided into groups 1< group 2< group 3< group 4< group 5, indicating that the academic performance of students has been gradually improved since the second semester of the third grade. The mean difference between group 1 and group 3 was 6.084*, indicating that the academic performance in the second semester of grade 4 was significantly higher than that in the second semester of grade 3. The mean difference between group 1 and group 4 was 8.315*, indicating that the academic performance in the first semester of grade 5 was significantly higher than that in the second semester of grade 3. The mean difference between group 1 and group 5 was 14.022*, indicating that the academic performance in the second semester of grade 5 was significantly higher than that in the second semester of grade 3. The mean difference between group 1 and group 2 was 2.230, indicating that the mean academic achievement in the first semester of grade four was higher than that in the second semester of grade three, but there was no significant difference.

It can be seen from Table 4 that the mean difference between group 2 and group 3 is 3.764*, p=0.002<0.05, indicating that the mean academic performance of the second semester of the fourth grade is significantly higher than that of the first semester of the fourth grade. The mean difference between group 3 and group 4 is 2.320, indicating that the mean academic achievement of the first semester of grade 5 is

higher than that of the second semester of grade 4, but there is no significant difference. The mean difference between group 4 and group 5 is 5.708*, indicating that the academic performance in the second semester of grade 5 is significantly higher than that in the first semester of grade 5.

4.2. Three years of tracking changes in attitudes to science learning

**Table 5**. Student science learning attitude achievement statistical description table

| | Least | Max | Mean | Variance | Skewness | | Kurtosis | |
|---|---|---|---|---|---|---|---|---|
| | | | | | S | E | S | E |
| 5-2 | 178 | 3.5 | 30.55 | 75.799 | -1.018 | .182 | .215 | 0.362 |
| 5-1 | 178 | 0 | 36.19 | 75.151 | -2.091 | .182 | 4.951 | 0.362 |
| 4-2 | 178 | 5 | 33.68 | 57.373 | -1.222 | .182 | 1.565 | 0.362 |
| 4-1 | 178 | 0 | 33.19 | 45.464 | -1.193 | .182 | 3.311 | 0.362 |
| 3-2 | 178 | 0 | 21.29 | 91.465 | -1.213 | .182 | .343 | 0.362 |

*Note:* 5-2: The second term of Grade Five,5-1: The first term of Grade Five,4-2: The second term of Grade Fourth,4-1: The first term of Grade Fourth,3-2: The second term of Grade Three, S: Statistics, E: Error of standard

According to the data in Table 5, the average of students' attitude towards science learning in the second semester of the third grade is 21.29, and that turned to 33.19,33.68,36.19 and 30.55 in the later 4 semesters. Generally speaking, the students' attitude towards science in the five semesters showed an upward trend, while their scores in the second semester of the fifth grade decreased, which may be mainly due to the impact of the epidemic closure of classes.

**Table 6**. ANOVA variance test of scientific learning attitudes

| | QS | DOF | MS | F | p |
|---|---|---|---|---|---|
| Interblock | 23455.998 | 4 | 5864.000 | 84.924 | .000 |
| Intra-class | 61109.549 | 885 | 69.050 | | |
| Total | 84565.547 | 889 | | | |

*Note:* QS: Quadratic sum,MS: Mean square,S:Significance

According to the results in Table 6, there are significant differences among the five groups of data, indicating that students' science learning attitude and achievement in the five semesters have significant differences, which can be furtherly compared after the fact.

**Table 7**. Table of results of post-test analysis of scientific learning attitude

| Dependent Variable: learning attitude | | | | | | |
|---|---|---|---|---|---|---|
| **HSD** | | | | | | |
| | | | | | 95% confidence interval | |
| (I) VAR00002 | (J) VAR00002 | MD(I-J) | SE | P | LL | UL |
| 1 | 2 | -11.902* | .877 | .000 | -14.374 | -9.429 |
| | 3 | -12.390* | .914 | .000 | -14.967 | -9.814 |
| | 4 | -14.742* | .968 | .000 | -17.467 | -12.016 |
| | 5 | -9.264* | .969 | .000 | -11.995 | -6.533 |
| 2 | 1 | 11.902* | .877 | .000 | 9.429 | 14.374 |
| | 3 | -.489 | .760 | .999 | -2.630 | 1.653 |
| | 4 | -2.840* | .823 | .006 | -5.160 | -.520 |
| | 5 | 2.638* | .825 | .015 | .312 | 4.964 |
| 3 | 1 | 12.390* | .914 | .000 | 9.814 | 14.967 |
| | 2 | .489 | .760 | .999 | -1.653 | 2.630 |
| | 4 | -2.351 | .863 | .066 | -4.782 | .080 |
| | 5 | 3.126* | .865 | .003 | .689 | 5.564 |
| 4 | 1 | 14.742* | .968 | .000 | 12.016 | 17.467 |
| | 2 | 2.840* | .823 | .006 | .520 | 5.160 |
| | 3 | 2.351 | .863 | .066 | -.080 | 4.782 |
| | 5 | 5.478* | .921 | .000 | 2.883 | 8.072 |
| 5 | 1 | 9.264* | .969 | .000 | 6.533 | 11.995 |
| | 2 | -2.638* | .825 | .015 | -4.964 | -.312 |
| | 3 | -3.126* | .865 | .003 | -5.564 | -.689 |
| | 4 | -5.478* | .921 | .000 | -8.072 | -2.883 |

*. The significance level of the mean difference is 0.05.

*Note:* MD: Mean difference, SE: Standard error, LL: lower limit, UL: upper limit

It can be seen from Table 7 that the scores of science learning attitude are divided into groups 1< group 2< group 3< group 4, indicating that the scores of students' science learning attitude are gradually increasing from the second semester of Grade 3 to the end of the first semester of grade 5. Group 4> group 5, which means that the achievement of science learning attitude has declined since the second semester of grade 5, and the reasons can be furtherly analyzed. The mean difference between group 1 and group 2 was 11.902*, indicating that the achievement of science learning attitude in the first semester of the fourth grade was significantly higher than the mean in the second semester of the third grade. The mean difference between group 1 and group 3 was 12.390*, indicating that the achievement of science learning attitude in the second semester of grade 4 was significantly higher than that in the second semester of grade 3. The mean difference between group 1 and group 4 was 14.742*, indicating that the achievement of science learning attitude in the first semester of grade 5 was significantly higher than that in the second semester of grade 3. The mean difference between group 1 and group 5 was 9.264*, indicating that the achievement of science learning attitude in the second semester of grade 5 was significantly higher than that in the second semester of grade 3.

As can be seen from Table 7, the mean difference between group 2 and group 3 is 0.489, p=0.999>0.05, indicating that the mean score of science learning attitude in the second semester of the fourth grade is higher than that in the first semester of the fourth grade, but there is no significant difference. The mean difference between group 3 and group 4 was 0.489, p=0.066>0.05, indicating that the mean academic achievement of the first semester of grade 5 was higher than that of the second semester of grade 4, but there was no significant difference. The mean difference between group 4 and group 5 was -5.478*, indicating that the academic performance of the second semester of grade 5 was significantly lower than that of the first semester of grade 5, which might be because the second semester of grade 5 was caused by the epidemic, and 2 months of the semester was online teaching, which can be further analyzed.

4.3. Changes in extracurricular performance, classroom performance and extended evaluation were tracked for three years

4.3.1. Changes in students' extracurricular performance

**Table 8**. Student extracurricular performance score statistical description table

| | Least | Max | Mean | Variance | Skewness | | Kurtosis | |
|---|---|---|---|---|---|---|---|---|
| | | | | | S | E | S | E |
| 5-2 | 0 | 19 | 12.90 | 16.284 | -1.547 | .182 | 2.302 | .362 |
| 5-1 | 0 | 20 | 17.25 | 15.642 | -2.507 | .182 | 6.637 | .362 |
| 4-2 | 0 | 20 | 17.47 | 15.292 | -2.785 | .182 | 8.386 | .362 |
| 4-1 | 0 | 20 | 18.14 | 15.125 | -2.943 | .182 | 9.260 | .362 |
| 3-2 | 0 | 20 | 14.22 | 46.540 | -.890 | .182 | -.396 | .362 |

*Note:* 5-2: The second term of Grade Five,5-1: The first term of Grade Five,4-2: The second term of Grade Fourth,4-1: The first term of Grade Fourth,3-2: The second term of Grade Three, S: Statistics, E: Error of standard

According to the data in Table 8, the average extracurricular performance score of students in the second semester of the third grade is 14.22, the average extracurricular performance score in the first semester of the fourth grade is 18.14, the average extracurricular performance score in the second semester of the fourth grade is 17.47, and the average extracurricular performance score in the first semester of the fifth grade is 17.25. At the end of the second semester of the fifth grade, the average extracurricular performance score was 12.90. Generally speaking, the extracurricular performance score of the experimental class showed a trend of increasing first and then decreasing.

**Table 9**. Results table of ANOVA analysis of variance

| | QS | DOF | MS | F | p |
|---|---|---|---|---|---|
| Interblock | 3757.548 | 4 | 939.387 | 43.138 | .000 |
| Intra-class | 19272.195 | 885 | 21.776 | | |
| Total | 23029.743 | 889 | | | |

*Note:* QS: Quadratic sum,MS: Mean square,S:Significance

Analysis of variance on students' extracurricular performance scores for five consecutive semesters, F=43.138, P<0.001, There are significant differences among the five groups of data, indicating that the final extracurricular performance of students in the five semesters has significant differences, which can be further compared after the event. In the spring of 2021, the second semester of the fourth grade, the extracurricular performance began to decline, and in the spring of 2022, the second semester of the fifth grade dropped significantly. In the spring of 2021 (the second semester of Grade four), classes were

suspended for 2 weeks due to the epidemic; in the autumn of 2021 (the first semester of grade five), classes were suspended once due to the epidemic and online teaching lasted for 1 month; in the spring of 2022 (the second semester of grade five), classes were suspended for 3 times due to the epidemic and online teaching lasted for 2 months in total. Online teaching has great influence on students to complete extracurricular science experiments and daily homework writing.

**Table 10**. After-class performance test results analysis table

| Dependent Variable: Extracurricular performance | | | | | | |
|---|---|---|---|---|---|---|
| | | HSD | | | | |
| | | | | | 95% confidence interval | |
| (I) VAR00002 | (J) VAR00002 | MD(I-J) | SE | P | LL | UL |
| 1 | 2 | -3.919* | .495 | .000 | -5.271 | -2.567 |
| | 3 | -3.250* | .495 | .000 | -4.602 | -1.898 |
| | 4 | -3.022* | .495 | .000 | -4.375 | -1.670 |
| | 5 | 1.329 | .495 | .057 | -.023 | 2.681 |
| 2 | 1 | 3.919* | .495 | .000 | 2.567 | 5.271 |
| | 3 | .669 | .495 | .659 | -.684 | 2.021 |
| | 4 | .896 | .495 | .367 | -.456 | 2.248 |
| | 5 | 5.247* | .495 | .000 | 3.895 | 6.599 |
| 3 | 1 | 3.250* | .495 | .000 | 1.898 | 4.602 |
| | 2 | -.669 | .495 | .659 | -2.021 | .684 |
| | 4 | .228 | .495 | .991 | -1.125 | 1.580 |
| | 5 | 4.579* | .495 | .000 | 3.227 | 5.931 |
| 4 | 1 | 3.022* | .495 | .000 | 1.670 | 4.375 |
| | 2 | -.896 | .495 | .367 | -2.248 | .456 |
| | 3 | -.228 | .495 | .991 | -1.580 | 1.125 |
| | 5 | 4.351* | .495 | .000 | 2.999 | 5.703 |
| 5 | 1 | -1.329 | .495 | .057 | -2.681 | .023 |
| | 2 | -5.247* | .495 | .000 | -6.599 | -3.895 |
| | 3 | -4.579* | .495 | .000 | -5.931 | -3.227 |
| | 4 | -4.351* | .495 | .000 | -5.703 | -2.999 |

*. The significance level of the mean difference is 0.05.

*Note:* MD: Mean difference, SE: Standard error, LL: lower limit, UL: upper limit

It can be seen from Table 10 that the mean difference between group 1 and group 2 is 3.919*, indicating that the extracurricular performance of the first semester of the fourth grade is significantly higher than that of the second semester of the third grade. The mean difference between group 1 and group 3 was 3.25*, indicating that the extracurricular performance of the second semester of the fourth grade was significantly higher than that of the second semester of the third grade. The mean difference between group 1 and group 4 was 3.022*, indicating that the extracurricular performance in the first semester of grade 5 was significantly higher than that in the second semester of grade 3. The mean difference between group 1 and group 5 was -1.329, p=0.057>0.05, indicating that the mean value of extracurricular activities in the second

semester of grade 5 was lower than that in the second semester of grade 3, but there was no significant difference.

It can be seen from Table 10 that the mean difference between group 2 and group 3 is -0.669, indicating that the scores of the second semester of the fourth grade are lower than the scores of the first semester of the fourth grade, with no significant difference. The mean difference between group 3 and group 4 is -0.228, indicating that the mean score of the first semester of grade 5 is lower than the mean score of the second semester of grade 4, but there is no significant difference. The mean difference between group 4 and group 5 was -4.315, indicating that the score of the second semester of grade 5 was significantly lower than that of the first semester of grade 5, which might be due to the epidemic situation in the second semester of grade 5. There were two months of online teaching.

4.3.2. Student performance score changes in class

**Table 11**. Student classroom performance score statistical description table

| | Least | Max | Mean | Variance | Skewness | | Kurtosis | |
|---|---|---|---|---|---|---|---|---|
| | | | | | S | E | S | E |
| 5-2 | 0 | 20 | 16.10 | 32.253 | −1.544 | .182 | 1.319 | .362 |
| 5-1 | 0 | 20 | 16.97 | 28.318 | −2.065 | .182 | 3.469 | .362 |
| 4-2 | 0 | 20 | 14.74 | 26.898 | −1.010 | .182 | .434 | .362 |
| 4-1 | 0 | 20 | 14.42 | 20.934 | −.820 | .182 | .375 | .362 |
| 3-2 | 0 | 20 | 6.85 | 11.866 | −.894 | .182 | −.493 | .362 |

*Note:* 5-2: The second term of Grade Five,5-1: The first term of Grade Five,4-2: The second term of Grade Fourth,4-1: The first term of Grade Fourth,3-2: The second term of Grade Three, S: Statistics,E: Error of standard.

According to the data in Table 11, the average classroom performance score of students in the second semester of the third grade is 6.85, the average final score of the first semester of the fourth grade is 14.42, the average final score of the second semester of the fourth grade is 14.74, the average final score of the first semester of the fifth grade is 16.97, and the average final score of the second semester of the fifth grade is 16.10. In general, the performance scores of students in the experimental class increased in the five semesters, but fluctuated in the second semester of the fifth grade. The main reason may be the influence of online teaching.

**Table 12**. Classroom performance scores ANOVA variance analysis results table

| | QS | DOF | MS | F | p |
|---|---|---|---|---|---|
| Interblock | 11542.538 | 4 | 2885.635 | 119.966 | .000 |
| Intra-class | 21287.551 | 885 | 24.054 | | |
| Total | 32830.089 | 889 | | | |

*Note: LS:* QS: Quadratic sum, MS: Mean square, S:Significance

Analysis of variance on students' performance scores in science classes for five consecutive semesters. F=119.966, P<0.001, There are significant differences between the five groups of data, indicating that the final scores of students in the five semesters have significant differences, which can be further compared after the event.

**Table 13**. Analysis table of post-test results of classroom performance

| Dependent Variable: classroom performance | | | | | | |
|---|---|---|---|---|---|---|
| HSD | | | | | | |
| | | | | | 95% confidence interval | |
| (I) VAR00002 | (J) VAR00002 | MD(I-J) | SE | P | LL | UL |
| 1 | 2 | -7.562* | .511 | .000 | -8.960 | -6.164 |
| | 3 | -7.888* | .511 | .000 | -9.285 | -6.490 |
| | 4 | -10.112* | .511 | .000 | -11.510 | -8.715 |
| | 5 | -10.112* | .511 | .000 | -11.510 | -8.715 |
| 2 | 1 | 7.561* | .511 | .000 | 6.164 | 8.959 |
| | 3 | -.326 | .511 | .969 | -1.723 | 1.072 |
| | 4 | -2.551* | .511 | .000 | -3.948 | -1.153 |
| | 5 | -2.551* | .511 | .000 | -3.948 | -1.153 |
| 3 | 1 | 7.888* | .511 | .000 | 6.490 | 9.285 |
| | 2 | .326 | .511 | .969 | -1.072 | 1.723 |
| | 4 | -2.224* | .511 | .000 | -3.622 | -.827 |
| | 5 | -2.224* | .511 | .000 | -3.622 | -.827 |
| 4 | 1 | 10.112* | .511 | .000 | 8.715 | 11.510 |
| | 2 | 2.551* | .511 | .000 | 1.153 | 3.948 |
| | 3 | 2.225* | .511 | .000 | .827 | 3.622 |
| | 5 | .862 | .511 | 0.460 | -1.398 | 1.398 |
| 5 | 1 | 10.112* | .511 | .000 | 8.715 | 11.510 |
| | 2 | 2.551* | .511 | .000 | 1.153 | 3.948 |
| | 3 | 2.225* | .511 | .000 | .827 | 3.622 |
| | 4 | -.862 | .511 | 0.460 | -1.398 | 1.398 |

*. The significance level of the mean difference is 0.05.

*Note:* MD: Mean difference, SE: Standard error, LL: lower limit, UL: upper limit

As can be seen from Table 13, class performance is divided into groups 1< group 2< group 3< group 4, indicating that class performance increases gradually from the second semester of Grade 3 to the end of the first semester of grade 5. Group 4> Group 5, indicating that the classroom performance has declined since the second semester of Grade 5, and the reasons can be further analyzed. The mean difference between group 1 and group 2 was 7.562*, indicating that the classroom performance of the first semester of the fourth grade was significantly higher than the mean value of the second semester of the third grade. The mean difference between group 1 and group 3 was 7.888*, indicating that the classroom performance of the second semester of the fourth grade was significantly higher than that of the second semester of the third grade. The mean difference between group 1 and group 4 was 10.112*, indicating that the classroom performance in the first semester of grade 5 was significantly higher than that in the second semester of grade 3. The mean difference between group 1 and group 5 was 10.112*, indicating that the classroom performance in the second semester of grade 5 was significantly higher than that in the second semester of grade 3.

As can be seen from Table 13, the mean difference between group 2 and group 3 is -0.326, p=0.969>0.05, indicating that the classroom performance of the second semester of the fourth grade is higher than that of the first semester of the fourth grade, but there is no significant difference. The mean difference between group 3 and group 4 is -2.225, indicating that the mean of classroom performance in the first semester of grade 5 is significantly higher than that in the second semester of grade 4. The mean difference between group 4 and group 5 was -0.862, indicating that the classroom performance in the second semester of grade 5 was lower than that in the first semester of grade 5, with no significant difference.

4.3.3. Changes in extensibility evaluation scores

**Table 14**. Statistical description table of extensibility evaluation score

| | Least | Max | Mean | Variance | Skewness | | Kurtosis | |
|---|---|---|---|---|---|---|---|---|
| | | | | | S | E | S | E |
| 5-2 | 0 | 7 | 1.56 | 2.723 | 1.056 | .182 | .542 | .362 |
| 5-1 | 0 | 10 | 1.82 | 4.849 | 1.336 | .182 | 1.762 | .362 |
| 4-2 | 0 | 10 | 1.47 | 2.521 | 1.556 | .182 | 4.215 | .362 |
| 4-1 | 0 | 10 | .63 | 1.894 | 3.667 | .182 | 17.512 | .362 |
| 3-2 | 0 | 6 | .21 | .564 | 4.802 | .182 | 27.193 | .362 |

*Note:* 5-2: The second term of Grade Five, 5-1: The first term of Grade Five, 4-2: The second term of Grade Fourth, 4-1: The first term of Grade Fourth, 3-2: The second term of Grade Three, S: Statistics, E: Error of standard.

According to the data in Table 14, the average score of the extended evaluation in the second semester of the third grade is 0.21, the average final score in the first semester of the fourth grade is 0.63, the average final score in the second semester of the fourth grade is 1.47, the average final score in the first semester of the fifth grade is 1.82, and the average final score in the second semester of the fifth grade is 1.56. In general, the scores of the experimental students in the five semesters showed an increasing trend, but there was a fluctuation in the second semester of the fifth grade. The main reason is that the online teaching for two months reduced the number of students participating in activities. Although online activities were held, the effect was not good.

**Table 15**. Results table of ANOVA analysis of variance for extensibility evaluation score

| | QS | DOF | MS | F | p |
|---|---|---|---|---|---|
| **Interblock** | 330.366 | 4 | 82.592 | 32.901 | .000 |
| **Intra-class** | 2221.635 | 885 | 2.510 | | |
| **Total** | 2552.001 | 889 | | | |

*Note: LS:* QS: Quadratic sum, MS: Mean square, S: Significance

Analysis of variance on students' scores in science extension assessment for five consecutive semesters, F=32.901, P<0.001. There are significant differences between the five groups of data, indicating that the final scores of students in the five semesters have significant differences, which can be further compared after the event.

**Table 16**. Analysis table of post-test results of extensibility evaluation

| Dependent Variable: extensibility evaluation | | | | | | |
|---|---|---|---|---|---|---|
| HSD | | | | | | |
| | | | | | 95% confidence interval | |
| (I) VAR00002 | (J) VAR00002 | MD(I-J) | SE | P | LL | UL |
| 1 | 2 | -.421 | .168 | .090 | -.880 | .038 |
| | 3 | -1.252* | .168 | .000 | -1.712 | -.793 |
| | 4 | -1.607* | .168 | .000 | -2.066 | -1.148 |
| | 5 | -1.343* | .168 | .000 | -1.802 | -.884 |
| 2 | 1 | .421 | .168 | .090 | -.0378 | .880 |
| | 3 | -.831* | .168 | .000 | -1.291 | -.372 |
| | 4 | -1.185* | .168 | .000 | -1.645 | -.726 |
| | 5 | -.921* | .168 | .000 | -1.380 | -.462 |
| 3 | 1 | 1.252* | .168 | .000 | .793 | 1.711 |
| | 2 | .831* | .168 | .000 | .372 | 1.291 |
| | 4 | -.354 | .168 | .218 | -.813 | .105 |
| | 5 | -.089 | .168 | .984 | -.549 | .369 |
| 4 | 1 | 1.607* | .168 | .000 | 1.148 | 2.066 |
| | 2 | 1.185* | .168 | .000 | .726 | 1.645 |
| | 3 | .354 | .168 | .218 | -.105 | .813 |
| | 5 | .264 | .168 | .516 | -.195 | .723 |
| 5 | 1 | 1.342* | .168 | .000 | .884 | 1.802 |
| | 2 | .921* | .168 | .000 | .462 | 1.380 |
| | 3 | .900 | .168 | .984 | -.369 | .549 |
| | 4 | -.264 | .168 | .516 | -.723 | .195 |

*. The significance level of the mean difference is 0.05.

*Note:* MD: Mean difference, SE: Standard error, LL: lower limit, UL: upper limit

It can be seen from Table 16 that the scores of the extended assessment are all divided into groups 1< group 2< group 3< group 4, indicating that the average scores of the extended assessment are gradually increasing from the second semester of the third grade to the end of the first semester of the fifth grade. Group 4> Group 5 shows, that the performance of extended evaluation has declined since the second semester of grade 5, and the reasons can be further analyzed. The mean difference between group 1 and group 3 was 1.253*, indicating that the extended performance in the second semester of grade 4 was significantly higher than that in the second semester of grade 3. The mean difference between group 1 and group 4 was 1.607*, indicating that the extended performance in the first semester of grade 5 was significantly higher than that in the second semester of grade 3. The mean difference between group 1 and group 5 was 1.343*, indicating that the extended performance in the second semester of grade 5 was significantly higher than that in the second semester of grade 3. The mean difference between group 1 and group 2 was 0.421*, p=0.90>0.05, indicating that the extended scores in the first semester of the fourth grade were higher than the mean scores in the second semester of the third grade, but there was no significant change.

It can be seen from Table 16 that the mean difference between group 2 and group 3 is 0.831*, indicating that the score of the second semester of the fourth grade is significantly higher than that of the first semester of the fourth grade. The mean difference between group 3 and group 4 is 0.354, indicating that the mean score of the first semester of grade 5 is higher than that of the second semester of grade 4, and there is no significant difference. The mean difference between group 4 and group 5 is -0.264, indicating that the scores of the second semester of Grade 5 are lower than the scores of the first semester of grade 5, and there is no significant difference.

**5. Discussion**

The effect of comprehensive evaluation on students' academic achievement in primary science should not be ignored. Comprehensive evaluation is the observation, record and analysis of students' overall development. (Xu, 2007) Evaluation is for promotion and evaluation is for development. (Hwang et al., 2018; Yang et al., 2021) This study found that comprehensive evaluation is conducive to the improvement of students' academic performance. Therefore, compared with a single paper-and-pencil test, comprehensive evaluation is conducive to the overall development of students.

The application of comprehensive evaluation can promote pupils' academic performance. Comprehensive evaluation has a great impact on the discussion and communication in group cooperation, the preparation and review of homework, and the attitude of participating in scientific activities. (Hwang et al., 2018; Yang et al., 2021) The most time-saving and effective way to organically integrate grade grading, learning and learning motivation is to ask not only teachers but also students to propose goals in the initial course design. Therefore, in the comprehensive evaluation of primary school science curriculum, it is necessary to consider students' classroom performance, extracurricular performance, extended evaluation and other factors, which will become one of the indicators affecting the comprehensive evaluation.

The application of comprehensive evaluation can promote students' attitude towards learning science. Guo et al. (2022) found that students with higher academic levels have more positive attitudes towards science. William, a British scholar, found through long-term classroom observation and investigation that it can stimulate learning motivation, provide feedback to promote learning, and encourage learners to become the master of their own learning. (Dylan Wiliam, 2019)

In Table 4 of this study, it is found that the scores of science learning attitude are grouped into 1< group 2< group 3< group 4, indicating that the scores of students' science learning attitude are gradually increasing from the second semester of Grade 3 to the end of the first semester of grade 5. Group 4> Group 5, indicating that the performance of scientific learning attitude began to decline from the second semester of grade 5, which may be due to the epidemic in the second semester of grade 5. In the second semester of grade 5, online teaching was used for 2 months, which could be further analyzed in the classroom performance, extracurricular performance and extended evaluation.

The application of comprehensive evaluation can promote students' classroom performance. Larraz N (2015) believes that cooperative learning can improve students' self-discipline, develop and improve teamwork, reflection and other aspects. Yalvema & Miaz (2017) Cooperation is very important to promote students' learning, and cooperative learning is better than single-work learning. (Chickering & Gamson, 1987) Cooperative learning can promote students' self-control in the classroom. (Perry et al.,1996) In Table 13 of this study, it is found that class performance is divided into groups 1< group 2< group 3< group 4, indicating that class performance increases gradually from the second semester of Grade 3 to the end of the first semester of grade 5. Group 4> Group 5, indicating a decline in classroom performance from the second semester of grade 5. The application of comprehensive evaluation can promote students' extended evaluation. Students' participation in science activities plays an important role in promoting students'

academic achievement in science (Rueggw, 2004). In Table 16 of this study, it is found that the scores of the extended assessment are all divided into groups 1< group 2< group 3< group 4, indicating that the average scores of the extended assessment gradually increase from the second semester of the third grade to the end of the first semester of the fifth grade. Group 4> Group 5, indicating that the performance of extended evaluation has declined since the second semester of grade 5. Therefore, it can be seen that the science curriculum should attach importance to students' independent participation or participation in various scientific activities organized by the school. This study found that students should be encouraged to actively participate in science and technology practice activities inside and outside the school in comprehensive evaluation of primary school science.

In the second semester of Grade 5, students' average scores in both classroom performance and extended evaluation showed a significant decline, which is mainly due to the impact of the local epidemic when online teaching was used for two months and students' classroom performance and students' participation in scientific activities were seriously limited.

The application of comprehensive evaluation has certain influence on students' extracurricular performance. Emsen & Perihan (2011) found that whether students' attitudes towards research assignments were significantly different depending on gender, class level and school type. From the second semester of the third grade to the first semester of the fourth grade, it is found in Table 10 of this study that the score increases and then decreases. In the spring of 2021, the second semester of the fourth grade, the extracurricular performance began to decline, and in the spring of 2022, the second semester of the fifth grade dropped significantly. In the spring of 2021 (the second semester of Grade four), classes were suspended for 2 weeks due to the epidemic; in the autumn of 2021 (the first semester of grade five), classes were suspended once due to the epidemic and online teaching lasted for 1 month; in the spring of 2022 (the second semester of grade five), classes were suspended for 3 times due to the epidemic and online teaching lasted for 2 months in total. Online teaching has great influence on students to complete extracurricular science experiments and daily homework writing. This factor is the main reason for the decline of students' attitude towards science learning in the second semester of Grade Five.

**6. Conclusions and Implications**

In primary science education, many evaluation methods have been proposed and proved effective in improving students' academic performance. Previous studies mainly focused on the evaluation of scientific knowledge, scientific thinking, attitude responsibility and scientific literacy, but there were few studies on the improvement of pupils' scientific academic performance and attitude by comprehensive evaluation. Therefore, on the basis of five years of practical exploration to construct a comprehensive evaluation of primary school science learning, this study analyzed the index data of 178 students from four classes over a period of three years and five semesters, and found that comprehensive evaluation could improve the academic achievement and attitude of primary school students in science. By tracking the data of grade 3, the research results show that (a) the use of comprehensive evaluation to improve pupils' academic achievement in science presents a gradual increasing trend, and there is a significant difference during the school period. (b) Using comprehensive evaluation to improve pupils' attitude towards science learning, showing an overall growth trend. (c) The use of comprehensive evaluation to improve students' classroom performance and expansive evaluation plays a promoting role, showing a gradually increasing trend, and there is a significant difference during the study period. (d) Comprehensive evaluation has a certain impact on the improvement of students' extracurricular performance. Therefore, in the comprehensive evaluation of primary school science, classroom performance, extracurricular performance, academic level, and extension evaluation should be included in the comprehensive evaluation system.

Theoretically speaking, this study uses the three-year data tracking results of comprehensive evaluation to show that comprehensive evaluation improves pupils' academic performance and attitude in science, and shows an increasing trend in classroom performance, extracurricular performance, academic performance and extended evaluation, with significant differences during school. Therefore, the results of the research indicate that comprehensive evaluation can improve students' academic achievement and attitude towards science. This provides theoretical support for the research evaluation model and expands the diversified forms of educational evaluation.

From the perspective of practice, in order to improve pupils' academic achievement and attitude in science, a comprehensive evaluation of primary school science learning is constructed. At the same time, the rational application of the comprehensive evaluation of primary school science learning in the evaluation of primary school science teaching is not only conducive to the innovative education process, conducive to the overall development of students, but also can adapt to the needs of other different disciplines. This type of evaluation is worth practicing in more schools and disciplines.

To improve students' academic performance and attitude in primary school science education is the pursuit of global science education in the 21st century. The conclusions of this study can provide some implications for science teachers from all over the world to improve students' academic performance and attitudes in science education. There are still many possibilities for indexes and weights of comprehensive evaluation. In science teaching, long-term observation, exploration and practice are also needed to work out an appropriate comprehensive evaluation model based on local conditions.

**7. Limitations and future research**

The development of students' cognition with the increase of grades cannot be controlled in the study, and index and weight of the comprehensive evaluation are not necessarily very reasonable. Besides, offline teaching was implemented in the era of Covid-19 epidemic, which must have had a certain impact on the study. Therefore, the applicability of comprehensive assessments of primary science learning to other participants is uncertain. It is necessary to further study the weight of indicators from different schools and students' cognitive level and comprehensive evaluation to explore development.

**Acknowledgements:** This research is from the project (Grant number SHSJ2022011) supported by the Life and Practice Education Committee of China Tao Xingzhi Research Institute of China.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References:**

Andrews, P, et al. (2014). PISA, TIMSS and Finnish Mathematics Teaching: An Enigma in Search of an Explanation. *Educational Studies in Mathematics*,87(1),7-26.https://doi.org/10.1007/s10649-014-9545-3

Ajzen, I. (2001). Nature and Operation of Attitudes. *Annual Review of Psychology*, 52(1), 27-58. https://doi.org/10.1146/annurev.psych.52.1.27

Barbara, E. Walvoord & Virginia, Johnson, Anderson. (1998). Effective Grading: A Tool for Learning and Assessment, *San Francisco: Jossey-Bass*,1-17.

Astin, A. W. (1999). Student involvement: a developmental theory for higher education. *Journal of college student development*, (5): 519-529.

Bian, X. C. (2017). College comprehensive evaluation enrollment reform: Evolution logic, mode selection and countermeasure analysis. *Research in Education*, (7), 108-115.

Chen, S. J.,Tan, F. (2015).Program for International Student Assessment (PISA) development status and future trend. *Comparative Education Research,* 37(07), 107-112. https://doi.org/1003-7667（2015）07-0107-06

Chi, S. H, Wang, Z., Liu, X. et al. (2017). Associations among Attitudes, Perceived Difficulty of Learning Science, Gender, Parents' Occupation and Students' Scientific Competencies. *International Journal of Science Education*, 39(16), 1-18.https://doi.org/10.1080/09500693.2017.1366675

Dylan Wiliam. (2019). Embedded formative evaluation. Shanghai: Shanghai Foreign Language Education Press, 51.

Emsen, P., Kilic, D.S., Soran, H. (2011). An examination of the attitudes of secondary school students towards research homeworks according to different variants. *Procedia Social and Behavioral Sciences*, 2574-2577. https://doi.org/10.1016/j.sbspro.2011.04.148

Gronick, W. S., Slowiaczek, M. L. (1994). Parents Involvement in Childrens Schooling: A multidimensional Conceptualization and Motivation Model. *Child Development*, (65),237-252.https://doi.org/10.2307/1131378

Grabau,L.J.,Trudel,L., & Ma, X. (2022). Associations between science dispositions and science literacy: A comparison of the United States and Canada. *International Journal of Science Education,* 44(9), 1440-1461. https://doi.org/10.1080/09500693.2022.2080887

Gardner,P.L.(1975). Attitudes to Science.*Studies in Science Education*,(2),1-41.

Guo,Sh.CH.,Liu.E.SH.,(2022). Current situation and characteristics of pupils' attitude towards science: A large-scale evaluation and comparison based on 277 schools in Zhengzhou. *Studies on science popularization*,17(1),42-51. https://doi.org/10.19293/j.cnki.1673-8357.2022.01.004

Hascher,T. Bischof,F (2000).Integrated and traditional homework in the elementary school - A comparison of students' achievement, stress and attitudes towards school. *Psychologie in erziehung und unterricht,*(47),252-265.

Hwang,W.Y.,Li,Y.H., & Shadiev, R. (2018). Exploring effects of discussion on visual attention, learning performance, and perceptions of students learning with str-support. *Computers & Education,* 116, 225-236. https://doi.org/10.1016/j.compedu.2017.09.015

Huang,D.CH.(2001). Research on the Comprehensive Quality evaluation System of college students. *Higher Education Research in Heilongjiang Province*,(5),37-40. https://doi.org/1003-2614(2001)05-0037-03

Jin,P.P.Liu,L.(2020). The dilemma and breakthrough strategy of comprehensive evaluation of enrollment fairness under the background of new college entrance examination. *Contemporary Education Forum*,(4),10-17. https://doi:10.13694/j.cnki.ddjylt.20200610.002

Jones L R,Wheeler G,Centurino.V.A.S. (2015).Timss 2015 Science Framework.9-10.http://timss andpirls.bc.edu.

Kind,P.Jones,K.,Barmby,P.(2007).Developing Attitudes towards Science Measures. *International Journal of Science Education*,29(7),871-893.https://doi.org/10.1080/09500690600909091

Ke,ZH.(2010). Complexity Analysis of examination evaluation system reform: A Case study of comprehensive quality evaluation policy. *Global Education Outlook*,(2),25-30.

Liu,W.H.(2018).Research on Quality evaluation of primary and secondary education. *Teaching and Management*,(10),118-122.

Lin,J.F.(2013). Problems and strategies in the implementation of integrated science curriculum. *Chinese Journal of Education*,(11):14-18.

Lee,J.,Grigg,W.S.,Dion,G.S.(2010).The Nation's Report Card:Mathematics 2007. *Education Statistics Quarterly*,(4),68.

Larraz N,Sandra V 6 zquez ,Liesa M.(2017). Transversal skills development through coop erative learning.*Training teachers for the future.on the Horizon*,25(2),85-95.

Li,X.Y.Wang,Y.(2016). A Study on the Comprehensive Evaluation of University Entrance Examination in Japan. *Contemporary education science*,(18),48-53.

Li,S.L.Wu,L.T.(2016). Analysis on the reform of comprehensive evaluation system of primary and secondary school students' academic quality. *Teaching and Management*,(9),76-79.

Ministry of Education P. R. China. (2021). Guidelines for quality evaluation of compulsory education.3-17. www.moe.gov.cn/srcsite/A06/s3321/202103/t20210317_520238.html

Ministry of Education of the People's Republic of China. (2021). Notice on the Outline of the Action Plan for Scientific Literacy of the whole People (2021-2035),6-25.

http://www.gov.cn/zhengce/content/ 2021-06/25/content_5620813.htm..

Ministry of Education of the People's Republic of China. (2022). *Curriculum Standards of Science*. Beijing. Beijing Normal University Publishing Group.

Mullis,I.V.S. & Martin,M.O.（Ed.）.(2018).*TIMSS 2015 Assessment Frameworks*.3-27.https：//timssandpirls.bc. edu/timss2015/frameworks.html.

Ministry of Education of the People's Republic of China. (2017). *Curriculum Standards of Science in elementary schools.* Beijing. Beijing Normal University Publishing Group.

Ministry of Education P. R. China. (2022). *Science Curriculum Standards for Middle School.* Beijing: Beijing Normal University Press.

Ministry of Education P. R. China. (2017). *Science Curriculum Standards for Middle School.* Beijing: Beijing Normal University Press.

Ministry of Education of the People's Republic of China.(2019). *Suggestions on strengthening and improving experimental teaching in primary and secondary schools*.4-10. http://www.moe.gov.cn/srcsite/A06/s3321/201911/t20191128_ 409958.html.

Mohammadpour E. A.(2013).Three-level Multilevel Analysis of Singaporean Eighth-graders Science Achievement. *Learning and Individual Differences*,(26),212-220. https://doi.org/10.1016/j.lindif.2012.12.005

NCES.(2019). *NAEP Science Assessment*.5-23. https://nces.ed.gov/nationsreportcard/science/2019-05-23.

NGSS Lead States.(2013).*Next generation science standards: For states, by states. The National Academies Press,* Washington, DC. https://doi.org/10.17226/18290

OECD. (2016). *PISA 2015 results (volume I) : Excellence and equity in education.* OECD Publishing. https://doi.org/10.1787/9789264266490-en

OECD.(2017).*PISA 2015 Assessment and Analytical Framework：Science，Reading，Mathematic，Financial Literacy and Collaborative Problem Solving (Revised Edition).* Paris:OECD Publishing.

OECD. (2019). *PISA 2018 results (volume I) : What students know and can do.* OECD Publishing. https://doi.org/10.1787/5f07c754-en

Pan,J.H.(2011).Problems and strategies in the implementation of integrated science curriculum. *Educational scientific research*,(9),69-72.

Patalle A,Cooper H,Robinson J C.Parent(2008). Involvement in homework:a research synthesis.*Review of Educational Research*.(4),1039-1101. https://doi.org/10.3102/0034654308325185

Pezdek.K.Berry,T.Renno,P.A.(2002).Children's mathematics ahievement the role of parents'perceptions and their involvement in homework.*Joumal of Educational Psychology*,94(4),771-77.https://doi.org/10.1037//0022-0663.94.4.771

Randy,E.Bennett.(2018).Educational Assessment: What to Watch in a Rapidly Changing World. *Educational Measurement：Issues and Practice*,37(4),7-15. https://doi.org/10.1111/emip.12231

Ruegg,W.(2004).*Universities in the nineteenth and early twentieth centuries（1800-1945）:a history of the university in Europe.*New York,Melbourne:Cambridge University Press, 47-52.

The Ministry of Education of the People's Republic of China.(2020). *The overall plan for deepening educational evaluation reform in the New era.*10-13. www.moe.gov.cn/jyb_xxgk/moe_1777/more_1778/202010/t20201013_494381.html

Tee O P.Subramaniam R.(2018). Comparative Study of Middle School Students'Attitudes towards Science：Rasch Analysis of Entire TIMSS 2011 Attitudinal Data for England，Singapore and the U. S.A. as well as Psychometric Properties of Attitudes Scale. *International Journal of Science Education*,40(3),268-290. https://doi.org/10.1080/09500693.2017.1413717

Tian,&Xiang,(2012). Review and prospect of the development of Formative evaluation in Japan. *Global Education Outlook*, 41(03):3-6+18.

Wang,R.B.Lin,J.(2021). Does Experimental Teaching affect Pupils' academic Achievement in Science?. *Journal of China Examinations*,(12).67-76. https://doi.org/10.19360/j.cnki.11-3303/g4.2021.12.009

Weng,L.L.Zhu,C.K.(2017). Analysis on the effect of the reform of "Trinity" comprehensive evaluation enrollment mode in Zhejiang Province. *Shanghai Education and Research Institute*,(6),9-15. https://doi:10.16194/j.cnki.31-1059/g4.2017.06.003

Wael Musalamani, Ruhizan Mohammad Yasin, Kamisah Osman.(2021).Comparison of school Based-cooperative problem based learning(SB-CPBL) and conventional teaching on students' attitudes towards science.*Journal of Baltic Science Education*,(20), No. 2.https://doi.org/10.33225/jbse/21.20.261

Wang,P.(2020). The subject identity of comprehensive quality Evaluation: Present situation, problems and countermeasures. *Chinese Journal of Education*,(9),38. https://doi.org/1002-4808 (2020) 09-0038-05

Wu,Q.,Zhang.K.,Liu.X.N.(2000). Research on Index system and method of comprehensive Quality Evaluation of college students. *Agricultural Education in China*,(3),41-42.

Wang,D.,Ren.H.Y.,Wang.J.(2019). Study on the effectiveness of comprehensive quality evaluation results. *Research on Educational Development*,(20),18-29. https://doi.org/10.14121/j.cnki.1008-3855.2019.20.005

Weinburgh M H. Gender，Ethnicity，and Grade(2000).*Level as Predictors of Middle School Students'Attitudes toward Science.* Georgia:Georgia State University.

Xu,H.(2007). On the role of science Curriculum teachers under the background of new curriculum. *Teaching and Management*,(4).

XU Min Kushner Benson SN，Mudrey-camino R etc.(2010).The relatonship between Parental involvement self reguhrd earning and reading achievement of fifth grades apath analysis using the ECLS-K database.*Socialpsy-chopgy of Educaton*,13(2),237-269.

Yang, X., Zhang, M., Kong, L., Wang, Q., & Hong, J. C. (2021). The effects of scientific self-efficacy and cognitive anxiety on science engagement with the "question-observation-doing-explanation" model during school disruption in covid-19 pandemic. *Journal of Science Education and Technology*, 30(3), 380-393. https://doi.org/10.1007/s10956-020-09877-x

Yehudit od-cohen5 Judith Hadari.(2015).Implementing Small Group Instruction among No vice Teachers in Their First Year of teaching.*Procedia_Social and Behavioral Sciences*,209.https://doi.org/10.1016/j.sbspro.2015.11.286

Yalvema Miaz .(2015).Improveing Students' Achievement of Social Scinece By Using Jigsaw cooperative Learning Model at Primary School.*IOSR Journal of Reserrch&Method in Education(IOSR-JRME),*5(4),1-6.

Yang,W.Z.(2015). Integrated Application of Hierarchical Teaching and Process Evaluation: A Case study of Information Technology in primary school.*Modern primary and secondary education*, 31(10):100-103.

Zhao,H.(2017). *Investigation of Cooperative Learning in Primary school Science classroom.* Chong Qing:ChongQing University.

Zhang,D.H& Tang,X.(2017).The influence of extracurricular activities on middle school students' science learning in China.Internterntional Journal of science education,(39),1381-1402.DOI10.1080/09500693.2017.1332797