

# Differentiated instruction in primary after-school science programs: A quasi-experimental study

Shu Zhao, Wei Liu, Yuxin Zuo, Grace Yue Qi & Wanshan Hu

To cite this article: Shu Zhao, Wei Liu, Yuxin Zuo, Grace Yue Qi & Wanshan Hu (01 Mar 2026): Differentiated instruction in primary after-school science programs: A quasi-experimental study, The Journal of Educational Research, DOI: [10.1080/00220671.2026.2634836](https://doi.org/10.1080/00220671.2026.2634836)

To link to this article: <https://doi.org/10.1080/00220671.2026.2634836>



© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC



Published online: 01 Mar 2026.



Submit your article to this journal [↗](#)



Article views: 142



View related articles [↗](#)



View Crossmark data [↗](#)

## Differentiated instruction in primary after-school science programs: A quasi-experimental study

Shu Zhao<sup>a</sup> , Wei Liu<sup>b</sup> , Yuxin Zuo<sup>a</sup> , Grace Yue Qi<sup>c</sup>  and Wanshan Hu<sup>a</sup> 

<sup>a</sup>College of Education, Shaanxi Normal University, Xi'an, China; <sup>b</sup>Faculty of Education, Beijing Normal University, Beijing, China; <sup>c</sup>School of Humanities Media and Creative Comm, Massey University, Auckland, New Zealand

### ABSTRACT

This study examined differences in student academic performance, self-efficacy, and learning engagement in the context of differentiated instruction (DI) in a cross-grade after-school science program, a setting promoted by China's "Double Reduction Policy". Using a quasi-experimental design, 129 primary students were assigned to DI (experimental) or conventional instruction (control) groups, with a pretest confirming equivalent baselines. The results showed that the DI group showed higher academic achievement (particularly in higher-order cognitive achievements) and learning engagement compared to the control group. Furthermore, although the post-test self-efficacy scores showed no significant differences, the student interviews revealed a contrasting pattern. Overall, the findings provide evidence on how DI operates in complex classroom settings and provide practical implications for teaching.

### ARTICLE HISTORY

Received 20 October 2025  
Revised 6 January 2026  
Accepted 17 February 2026

### KEYWORDS

Differentiated instruction (DI); after-school programs; science education; primary-school cross-grade levels

### Introduction

After-school programs have gained much popularity in K-12 schools on a global horizon. Families and communities value after-school programs to support young learners and teens for their academic and personal development. A wide variety of after-school activities in the programs have been implemented in many countries to provide additional learning time and meet diverse learning needs (Boggs, 2024; Magaji et al., 2022; McMahan et al., 2021; Shea et al., 2023). Countries such as Sweden and Australia have implemented structured after-school initiatives to improve academic outcomes and promote physical well-being, respectively (Karlsudd, 2023; Veldman et al., 2020). In 2021, China introduced Double Reduction Policy to establish some essential principles for after-school services, which has significantly shaped the development of after-school programs. While Double Reduction Policy emphasizes the reduction of students' homework and off-campus tutoring load, the provision of extended-hour educational programs offered by primary schools is allowed and encouraged. As a result, primary schools have since launched different kinds of after-school interest programs. While such initiative is welcomed by parents, schools and teachers have experienced additional pressure due to the lack of sufficient funding and teacher participation. Many primary schools only afford to cross-grade levels after-school programs (Liu & Yuan, 2024). Although the programs are generally flexible and less formal than regular classroom instruction, this flexibility primarily allows students to choose whether to participate based on

personal interests at the semester begin, often spanning cross-grade levels. The non-formal nature emphasizes interest-driven learning rather than strictly adhering to a regulated course with predetermined objectives and assessments. However, once enrolled, students are required to attend classes at fixed times and locations, and teachers follow a planned instructional schedule while guiding learning and fostering student interests. While this structure accommodates students' preferences, it still does not fully achieve programs to tailor instruction to students' prior academic levels, especially in cross-grade settings.

Differentiated Instruction (DI) allow teachers to address the diverse needs of students by tailoring content, process, product and learning environment. It has been regarded as a significant approach for addressing educational challenges (Tomlinson, 2000). As an onto-epistemological approach for teaching, DI acknowledges that all learners possess the right to be appropriately challenged and to achieve their maximum potential within educational settings (Goodnough, 2010). In light of the increasing diversity encountered in contemporary classrooms, DI has attracted considerable interest in both research and teaching practice. At present, most of the literature on the adoption of DI has focused on K-12 settings, especially primary schools. However, little has focused on cross-grade students learning in after-school classes. Our study aims to fill the gap by examining the use of DI on the learning performance of students at different grade levels in one after-school science program as well as the challenges and opportunities involved in designing and implementing this DI inspired program in China.

**CONTACT** Grace Yue Qi  g.qi@massey.ac.nz  School of Humanities Media and Creative Comm, Massey University, Auckland, New Zealand.

© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

## Literature review

### *DI in K12 education*

DI is a student-centered approach that adjusts content, process, product, and learning environment based on learners' cognitive levels, interests, and learning styles. Rather than following a fixed model, DI reflects a flexible mindset that values student diversity (Tomlinson, 2000; Tomlinson & McTighe, 2006). Effective implementation of DI requires teachers to adapt their instruction according to student characteristics (Grecu, 2023), yet this is often challenging due to the complex skills and planning involved (Hattie, 2008; van Geel et al., 2019). Suprayogi and Valcke (2016) highlight five core features of DI: addressing learner differences, applying specific strategies, diversifying activities, monitoring individual needs, and maximizing outcomes. These practices depend on a clear understanding of student profiles through pre-assessment (Suwastini et al., 2021; Tomlinson & Moon, 2013). Based on such student profiles, specific strategies such as flexible grouping and tiered activities have been shown to enhance engagement and better accommodate learning differences (Ariss, 2017). In addition, providing differentiated learning materials—such as leveled texts, multimedia resources—plays a crucial role in supporting DI. These materials can be adapted in terms of complexity, modality, and presentation format to align with students' needs (Dack & Tomlinson, 2015; Subban, 2006; Tomlinson, 2014). Another widely used pedagogical approach for implementing DI is inquiry-based learning (IBL). Different forms of inquiry—structured, guided, and open, can be matched to students' developmental levels and autonomy, making IBL highly adaptable for differentiation (Banchi & Bell, 2008). For instance, lower level learners may benefit more from structured inquiry supported by explicit guidance, whereas high level students often thrive in open inquiry environments where they assume full responsibility for the investigative process. Recently, digital tools such as intelligent learning platforms have shown promise in supporting DI by enabling teachers to deliver tailored learning materials and by capturing real-time data on students' learning (Pozas et al., 2020).

Building on these theoretical foundations and classroom approaches, empirical studies have increasingly examined the effectiveness of DI, highlighting its potential to improve students' academic performance, motivation, and engagement—although findings remain mixed. For instance, Lai et al. (2020) explored the effect of DI on the learning performance of a group of Grade Six students in mathematics. They found that DI significantly improved students' mathematics self-efficacy, learning motivations, and problem-solving skills. In a randomized experiment, Haelermans et al. (2015) revealed a positive and significant effect of digital differentiation on the academic performance of Grade Eight students. Focusing on senior high-school classes, Pablico et al. (2017) explored the effects of DI implementation on science students using mixed methods research. The study was conducted with students in the same grade and found that DI was positively perceived by both teachers and students,

leading to improved student engagement and enhanced classroom dynamics. However, despite these promising findings, not all research supports the effectiveness of DI. Earlier meta-analyses such as Horak (1981) report minimal or even negative effects, suggesting inconsistencies in outcomes.

Current research on DI has primarily concentrated on its implementation within formal science classroom settings, leaving its applicability in informal learning environments—such as after-school programs—largely underexplored and insufficiently validated. Furthermore, much of the current literature concentrates on single-grade classrooms, with limited attention paid to cross-grade instructional environments where the complexity of applying DI may be even greater (Smale-Jacobse et al., 2019).

### *Science education in after-school programs*

As an extension of and complement to the classroom, after-school programs offers significant advantages by providing a more relaxed and IBL environment. It bridges the gap between classroom education and students' natural curiosity. This flexibility allows for a personalized IBL experience, which has been shown to increase motivation, engagement, and academic achievement, making after-school programs an important part of education reform (Vandell et al., 2007).

Regarding the science education in after-school program, from the perspective of Kolb's Experiential Learning Theory (1984), this program that allows students to construct knowledge through inquiry engagement and reflection can support individualized learning paths that adapt to each student's interest, and cognitive level. Previous researches have demonstrated the value of after-school science program. For example, research by Moreno et al. (2016) demonstrates that after-school STEM activities increase Grade Five students' STEM-related content knowledge and skills. Sahin et al. (2014) explored the characteristics of STEM-related after-school activities and found that these programs emphasize open-ended, collaborative scientific investigations. Such activities provide students with opportunities to develop 21st century skills while fostering interest and engagement in STEM fields. The findings of Newell et al. (2015) suggest that after-school programs that actively engage students in enjoyable and purposeful scientific activities can improve their preparedness for future science coursework and careers.

Since 2020, primary schools across China have increasingly implemented after-school programs that span multiple grade levels, with science education emerging as a popular offering. However, these cross-grade settings present significant challenges for educators, as students often exhibit wide variations in cognitive development, prior knowledge, and individual learning needs. While after-school science education offers distinct advantages—such as greater instructional flexibility and opportunities for IBL, ensuring its effectiveness in highly heterogeneous classrooms remains a pressing concern. In particular, identifying pedagogical strategies that can provide effective guidance and accommodate diverse learners in such contexts has become a critical issue.

Addressing this challenge serves as both the starting point and the practical foundation of the present study.

### ***DI in after-school science program***

Nowadays, after-school science programs have gained increasing global attention as an important complement to formal schooling, with countries such as the United States (Allen et al., 2019), the United Kingdom (Magaji et al., 2022), and Australia (Hayes, 2018) widely implementing after-school STEM or science enrichment initiatives. Despite their growing popularity, student diversity remains a central challenge, necessitating flexible program designs and targeted support to ensure equitable participation and meaningful learning experiences for all students.

Although existing literature indicates that DI—which integrates instructional design with students' learning preferences—holds promise for enhancing extracurricular science education, empirical findings remain inconsistent (Pablico et al., 2017). These inconsistencies can be attributed to substantial variations in research design, measurement tools, and instructional implementation across studies, which in turn lead to divergent conclusions regarding DI's effectiveness. Furthermore, many studies suffer from methodological limitations, such as small sample sizes and inadequate control of confounding variables, thereby undermining the reliability and generalizability of their findings (Smale-Jacobse et al., 2019). Most notably, there remains a distinct lack of systematic, large-scale research focused specifically on the application of DI within after-school science programs at the primary school cross-grade level—highlighting a critical gap that this study aims to address.

In China, after-school science programs creating a more favorable environment for individualized instruction. This advantage is further enhanced by the rapid development of intelligent educational technologies in both software and hardware, which enable teachers to collect and analyze real-time student data more efficiently (Durlak & Weissberg, 2007). As a result, educators are better equipped to implement evidence-based teaching practices, monitor student progress closely, and tailor instructional strategies to meet individual learning needs. These conditions collectively create a practical foundation for applying DI in after-school science education, allowing teachers to respond more effectively to diverse student profiles and promote deeper, more personalized learning (Maeng, 2017).

However, due to the pronounced heterogeneity among cross-grade students in after-school science programs—particularly in terms of cognitive development, prior knowledge, and learning needs—the design and evaluation of instruction in such settings become significantly more complex and challenging. These contextual factors complicate the implementation of pedagogical approaches and the measurement of learning outcomes, thereby highlighting the urgent need for context-sensitive, empirical research to explore effective instructional models within local educational environments.

The use of DI in after school cross-grade science programs is expected to reflect the respective advantages of

both instructional design that values student diversity and the flexible learning environment inherent in informal education. This study aims to examine how DI is associated with student learning performance, including their academic achievements, self-efficacy, and learning engagement, in a cross-grade after school science program. The following research questions are proposed:

1. How is DI associated with academic achievement in a cross-grade level after-school science program?
2. How is DI associated with self-efficacy in a cross-grade level after-school science program?
3. How is DI associated with learning engagement in a cross-grade level after-school science program?

## **Methods**

### ***Participants***

This study was conducted at an urban primary school in Western China and involved two cross-grade after-school science groups comprising a total of 129 students from Grades Three, Four and Five. At the start of the semester, these students voluntarily selected the science program from all after-school programs offered by the school based on their personal interests, at the same time, chose one of the two available class time slots. As a result, students enrolled voluntarily into the two classes. After enrollment was finalized, one class was randomly designated as the experimental group and the other as the control group. The experimental group, which met on Tuesdays, included 21 Grade Three, 22 Grade Four, and 21 Grade Five students (35 boys and 29 girls; average age = 11.20 years). The control group, which met on Wednesdays, comprised 22 Grade Three, 21 Grade Four, and 22 Grade Five students (36 boys and 29 girls; average age = 11.22 years). Prior to the experiment, informed consent was obtained from teachers, parents, and students through a formal consent process approved by the local Ethics Committee.

### ***Research design and procedures***

This study employed a quasi-experimental research design to explore the impact of DI on students' academic achievement, self-efficacy and learning engagement. Before the experiment, all students completed a self-efficacy questionnaire and Pretest of academic levels. Pretest results indicated no statistically significant differences between the experimental and control groups in terms of academic performance or self-efficacy, confirming group equivalence at baseline. During the eight weeks, DI was implemented in the experimental group, while the control group received traditional teaching approach and was scheduled to receive the same DI intervention after the completion of the post-test and interview (delayed intervention design). Both classes were taught by the same teacher with strong ICT skill. The teacher had no prior familiarity with the students and had undergone a two-month DI training provided by the research team prior to the commencement of the experiment. The training

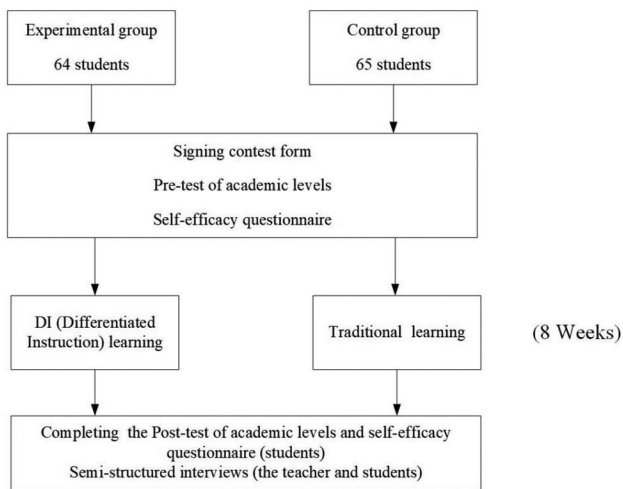


Figure 1. Research procedure.

Table 1. After-school science class topic and IBL.

Week	Topic	IBL
1	Sound	Inquiring the sound transmission effects of different objects.
2	Fermentation of yeast	Inquiring the effect of temperature on fermentation using yeast as the research object.
3	Heat transfer	Inquiring the contact between objects at different temperatures.
4	Heat transfer	Inquiring the transfer of heat in metals
5	Dissolve	Inquiring which factors can change the rate of disappearance of substances in water.
6	Magnet	Inquiring the factors that affect the proximity and bounce of two magnets toward each other.
7	Light	Inquiring why the size and shape of shadows change.
8	Oxidize	Inquiring the reasons for the yellowing of apples as the research object.

content includes DI concepts, task design, dynamic grouping, and the use of intelligent learning platforms.

The experiment was conducted in the physical classroom of the school. The classroom was equipped with an interactive whiteboard and 68 tablets. The students in the experimental group used the intelligent learning platform on the tablet to receive learning tasks and materials from the teacher to support learning. At the same time, the platform supports the teacher to deliver different resources to different student groups and helps teacher to collect the number of times every student accesses the learning materials, the duration of each learning, the time and accuracy required to complete tasks.

To mitigate potential teacher-induced bias, two control measures were implemented during the experiment. First, the teacher was required to participate in preparatory activities before each class that emphasized pedagogical neutrality and explicitly prohibited any intentional variation in instructional quality between the experimental and control groups. Second, after obtaining informed consent from both the teacher and the students, the research team conducted periodic reviews of classroom video recordings and systematically monitored students' activities on the learning platform, thereby ensuring that any observed differences could be attributed to the DI intervention rather than teacher-related bias.

The experiment was conducted eight weeks, and students participated in the after-school science program once a

Table 2. Differentiated learning activities design.

Student academic level	Content (learning materials)	Process (inquiry-based learning & activity)	Product (required outputs)
Low	Comprehensive materials: Including instructional videos (scientific knowledge, prior knowledge, detailed procedures) and a guided worksheet (text form) for documenting experimental results.	Structured Inquiry: Replicating the inquiry procedures demonstrated in the instructional videos.	Completing the teacher-provided guided worksheets (documenting results).
Medium	Focused materials: Including topic knowledge in a mind map, key procedural steps (text), and a worksheet for recording results and providing explanations.	Guided Inquiry: Following the key procedural steps to complete the inquiry.	Interpreting their inquiry results based on the worksheet prompts (recording results and providing explanations).
High	Minimal core materials: Including the core inquiry questions and a set of extended, higher-level materials (textual form).	Open Inquiry: Independently designing and carrying out their own inquiry procedures.	Producing independent inquiry reports accompanied by higher-order explanations (e.g., critical reflections and extended material synthesis).

week, each time lasting 90 minutes (see Figure 1). The weekly courses follow a structured sequence, consisting of three stages: presenting class topic, group IBL, and summative reflection. The class topic is shown in Table 1. In the experimental group, students were initially categorized into three academic levels—low (Grade Three), medium (Grade Four), and high (Grade Five).

Following Tomlinson's (2014) DI principles, differentiation in content, process, product, and learning environment were incorporated into the weekly courses to maximize students' learning potential, as shown in Table 2. Furthermore, learning environment differentiation was achieved through flexible and dynamic grouping and peer support. After each weekly course, the teacher evaluated student performance using data collected from the intelligent learning platform (including materials usage, learning duration, task time, and accuracy) to adjust their academic level and group for the subsequent week. Students exceeding the performance of their peers within the same level were promoted to a higher academic level group, while underperforming students were reassigned to a lower level group. Additionally, higher performing students in the highest academic level were designated as peer tutors to support low academic level peers throughout the inquiry process.

Meanwhile, in the control group, students were divided into three academic levels based on their grade and was also participated in presenting class topic, 4–5 students group IBL and summative reflection, but no differentiated in content, process, product, and learning environment were

**Table 3.** Overview of instruments, collected data, and analytical methods addressing each research question.

Research question	Instruments	Collected data	Analytical method
1. How is DI associated with academic achievement in a cross-grade level after-school science program?	Pretest of academic levels.	Pretest scores of the two groups' academic performance.	Mann-Whitney U test, Paired-samples t-test.
	Post-test of academic levels.	Post-test scores of the two groups' academic performance.	
	Outline of semi-structured interview for science teacher.	The teacher's perception of students' learning engagement.	Thematic analysis.
2. How is DI associated with self-efficacy in a cross-grade level after-school science program?	Self-efficacy questionnaire.	Self-efficacy scores of the two groups before and after the experiment.	Mann-Whitney U test.
	Outline of semi-structured interview for students.	The students' experience and self-efficacy.	Thematic analysis.
3. How is DI associated with learning engagement in a cross-grade level after-school science program?	Information Technology-based Interaction Analysis System (ITIAS).	Percentages of each variables (teacher talk, student talk, silence, technology) in every courses.	Mann-Whitney U test.
	Outline of semi-structured interview for science teacher.	The teacher's perception of students' performance.	Thematic analysis.

performed during the intervention phase. Specifically, although the control group used the same intelligent learning platform as the experimental group, the platform was used solely to provide identical learning materials to all students and was not employed to evaluate student performance for flexible or dynamic grouping or peer support. All students in the control group engaged in uniform activities and were required to produce the same learning products regardless of their academic levels. After eight weeks, the control group was provided with the same DI intervention following the procedures used in the experimental group.

Before and after the experiment, students from both the control and experimental groups completed the academic achievement assessment and self-efficacy questionnaire. Semi-structured interviews were also conducted with the teacher and the student participants to collect the data indicating their engagement process and experience.

### **Instruments and Data Analysis**

In this study, each instrument and analytical method was selected in accordance with the type of data collected and the specific aspect of the research questions it was intended to address. As the study examined three related yet distinct constructs, academic achievement, self-efficacy, and learning

engagement, multiple complementary data sources were required, including academic performance test scores, self-efficacy scores, percentages of each variable (teacher talk, student talk, silence, technology) in every course and student and teacher interviews. Accordingly, a combination of quantitative and qualitative analyses was employed. Table 3 summarizes the alignment between each instrument, the data it collected, and the analytical methods employed, illustrating how the different data worked together to provide a comprehensive and triangulated understanding of how DI was associated with academic achievement, self-efficacy and learning engagement.

### **Pre- and post-test of academic levels**

Guided by the PISA 2025 Science Framework (OECD, 2023), which emphasizes the dimensions of scientific knowledge and science competencies, the research team collaborated with the program's science teacher to design pre- and post-tests aimed at assessing students' academic achievement. The test design was grounded in Bloom's revised cognitive taxonomy (Anderson et al., 2000; Bloom, 1956), which categorizes cognitive processes into six hierarchical levels: remember, understand, apply, analyze, evaluate, and create. The first two levels—remember and understand—represent lower-order thinking skills (LOTS) and align with the dimension of scientific knowledge, while the latter four levels—apply, analyze, evaluate, and create—represent higher-order thinking skills (HOTS), corresponding to science competencies. The pre- and post-tests were parallel in structure and difficulty but contained different items to avoid practice effects. LOTS were assessed using fill-in-the-blank questions. For example, a pretest item asked, "The faster an object vibrates, the \_\_\_\_\_ of the sound changes," while a post-test item asked, "When the material of a sounding object differs, the \_\_\_\_\_ of the sound changes." HOTS were assessed through short-answer questions. A pretest example was, "Why do your ears feel painful when firecrackers explode during the Spring Festival?" while a post-test item asked, "An explosion at a factory produced a loud sound, and then cracks appeared in nearby windows. Why did this happen?" To validate the assessments, the tests were piloted with 72 students from a cross-grade science class in another local primary school, yielding 72 valid responses. Statistical analysis using SPSS 20.0 indicated high internal consistency, with Cronbach's alpha coefficients of 0.923 for the pretest and 0.889 for the post-test, confirming the reliability of the instruments for measuring academic outcomes.

Before the experiment, to ensure that the experimental and control groups did not differ statistically in their academic performance, the pretest of LOTS, HOTS, and total scores between the two groups were compared. After the experiment, examine the associated of DI with students' academic achievement, the post-test of LOTS, HOTS, and total scores between the two groups were compared. As the pre- and post-test data violated the assumptions of normality, and the residuals exhibited substantial skewness, parametric analyses such as ANOVA, independent-samples t-tests, or

ANCOVA were not appropriate. Therefore, the Mann-Whitney U test was employed as a more suitable non-parametric alternative. In addition, to further investigate which grade benefited the most from DI in terms of academic achievement, we conducted a paired-samples t-tests on the pre- and post-test scores within the experimental group with third, fourth and fifth grades. All of the students finished the pre- and post-tests. The test scores were evaluated by two teachers: the teacher of this program and a science teacher at another primary school. The Kendall's correlation coefficient for the total scores assigned by each teacher was .803, demonstrating significance at the .01 level. The final score for each test was computed based on the average score of two teachers. These processes collectively addressed the research question concerning the association with DI on students' academic performance.

### Self-efficacy questionnaire

The self-efficacy questionnaire administered to the students was developed by Pintrich and De Groot (1990) (see Table 4). The questionnaire included nine items, each of which was designed on a 5-point Likert scale ranging from 1, indicating never, to 5, indicating always. To ensure the content validity, the questionnaire was sent out to two experts to review. One expert, an English teacher from a tertiary institute, assessed the consistency of the English-Chinese translation and the overall comprehensibility of the questionnaire. The other expert, a primary school science teacher, provided feedback on the feasibility of the questionnaire, focusing on the length, formatting, and comprehensibility for primary students. The revised questionnaire was piloted in a cross-grade level science class at another local primary school, with the participation of 72 students who provided 72 valid responses. The reliability of the questionnaire was assessed using SPSS 20.0, and the results indicated that the Cronbach's Alpha coefficient for the questionnaire reached 0.841, demonstrating a high level of reliability.

All students in the experimental and control groups filled out the self-efficacy questionnaire at the same time that they completed the pre- and post-tests of academic level. As the self-efficacy scores of the two groups, both before and after the experiment, violated the assumption of normality and

the residuals were highly skewed, the 129 pretest questionnaires from the experimental and control groups were analyzed using the Mann-Whitney U test to confirm that there were no significant differences in self-efficacy between the groups prior to the experiment. After the experiment, the Mann-Whitney U test was again applied to the 129 post-test questionnaires. This procedure allowed to determine whether the self-efficacy of the two groups remained comparable after the experiment, thereby addressing the research question concerning the association of DI with students' self-efficacy.

### Information Technology-based Interaction Analysis System (ITIAS)

We utilized the Information Technology-based Interaction Analysis System (ITIAS), as shown in Table 5, to analyze the video data (Gu & Wang, 2004). ITIAS is an enhancement of the Flanders Interaction Analysis System (FIAS), the latter of which is used to quantify the verbal interactions between teachers and students in classroom teaching. FIAS categorizes teacher-student verbal behaviors into four major variables: teacher language, student language, silence, and disruption behaviors (Flanders, 1970). However, it does not reflect the use of information technology in the classroom or the interactions among students. Gu and Wang improved this system, proposing an enhanced version known as ITIAS. ITIAS includes additional behaviors related to student interactions and the use of technology in the classroom, making it more suitable for contemporary technology-enhanced classrooms and student-centered teaching trends. In this study, ITIAS was used as an observation tool, coupled with the ITIAS coding system, to encode and analyze course recordings. The goal was to understand differences in student learning engagement between the experimental and control groups.

After the consent of the teachers and students was obtained, we recorded the whole 16 courses, including both the experimental and control groups. All data in the class videos were analyzed. The researchers recorded the videos every three seconds, identifying the behaviors and coding them based on ITIAS. Initially, two researchers independently analyzed the same video using ITIAS as the coding framework. Intercoder reliability was assessed using Cohen's Kappa, yielding a coefficient of 0.805, indicating a substantial level of consensus between the coders. Subsequently, each researcher analyzed half of the remaining video. Finally, the researchers tallied the percentages of each variable in every course of the two groups. The Mann-Whitney U test was employed as the dataset consisted of two independent groups and the proportional data did not meet the assumption of normality. This analysis provided evidence to address the research question concerning the association with DI with students' learning engagement.

### Outline of semi-structured interviews for the science teacher and students

After the experimental phase concluded, a semi-structured interview was conducted with the science teacher (as an

Table 4. Self-efficacy questionnaire.

Question number	Statement
Q1	Compared with other students in the after-school science program, I expect to do well.
Q2	I'm certain that in the future, I can understand the ideas taught in the program.
Q3	I expect to do very well in this class.
Q4	Compared with others in the science class, I think I'm a good student.
Q5	I am sure I can do an excellent job on the problems and tasks assigned for this science class.
Q6	I believe I can receive positive reviews from the teacher and other classmates in the science club.
Q7	My study skills are excellent compared with others in this class.
Q8	Compared with other students in this science class, I think I have substantial knowledge about science.
Q9	I know that I will be able to learn more science knowledge for this class.

**Table 5.** Categories of Information Technology-based Interaction Analysis System (ITIAS).

Categories	Code	Description
Teacher talk	Indirect impact	1 Accepts feeling Accepts and clarifies the feeling tone of the students in a non-threatening manner. Feelings may be positive or negative. Predicting or recalling feelings are included.
		2 Praises or encourages Praises or encourages student action or behavior. Jokes that release tension, not at the expense of another individual, nodding head or saying, "Um-hm?" or "Go on" are included.
		3 Accepts or uses ideas of student Clarifying, building, or developing ideas suggested by a student. As the teacher brings more of the teacher's own ideas into play.
	Direct impact	4 Asks open questions Ask students questions based on the teacher's opinions or ideas and expect answers from students.
		5 Asks closed questions
		6 Lecturing Giving facts or opinions about content or procedures; Expressing the teacher's own ideas, asking rhetorical questions.
		7 Giving directions Directions, commands, or orders with which a student is expected to comply.
		8 Criticizing or justifying authority Statements intended to change student behavior from nonacceptable to acceptable patterns; bawling someone out; stating why the teacher is doing what; extreme self-reference.
Student Talk	9 Response (passive reaction) (Response to Code 4) Talk by students in response to teacher. Teacher initiates the contact or solicits student statement. Students' freedom to express themselves is limited.	
	10 Initiation Talk by students which they initiate. If "calling on" student is only to indicate who may talk next, observer must decide whether student wanted to talk. If he did, use this category.	
	11 Active questioning Taking the initiative to ask questions and expressing their opinions freely.	
	12 Discussion with the partner Discussion, exchange of views.	
Silence	13 Chaos which is not conducive to the teaching Temporary pauses, short periods of quiet or confusion that prevent observers from understanding student-teacher communication.	
	14 Reflect on the problem Students think about the problem.	
Technology	15 Practice Students do classroom exercises.	
	16 Teachers' controlling the technology Teachers use technology to present content and illustrate ideas	
	17 Students' controlling the technology Students use technology to present content and illustrate ideas; Students do classroom experiments	
	18 Technology works for students Students observe media presentations	

expert) who facilitated both student groups in the science program to explore the teacher's perceptions of student learning engagement and academic performance. The interview lasted approximately 20 minutes and focused on typical instructional situations, particularly group IBL. This situation was selected because it represented the key stage in which DI in the experimental group was most evident, specifically through differentiated content, process, product, and learning environment, all of which were designed to promote students' academic achievement and learning engagement. The purpose of this interview was to triangulate the findings from the pre- and post-test of academic levels of the two groups and the recorded in-class videos, and to provide explanatory insights where needed, thereby addressing the associations between DI and students' academic achievement and learning engagement in a cross-grade-level after-school science program. The interview questions for the teacher included the following, such as: (a) What do you think of the performance of students in each group? (b) What do you think of the learning engagement of students in each group?

Additionally, 10 students from the experimental group (3 Grade Three, 4 Grade Four, and 3 Grade Five) and 10 students from the control group (3 Grade Three, 4 Grade Four, and 3 Grade Five) were randomly selected to form two focus groups. Each focus group interview lasted about 30 minutes and addressed three typical instructional situations (presenting class topic, group IBL, and summative reflection), allowing students to describe their experience and self-efficacy in the science program. The interview questions included the following, such as: (a) What do you think about the class? (b) How do you evaluate your performance in the class? The data from these interviews were analyzed

alongside the self-efficacy questionnaire results to examine the association between DI and the self-efficacy of students in a cross-grade level after-school science program.

All interviews were audio-recorded and transcribed verbatim. The evaluation of the interviews followed the procedures proposed by Creswell and Clark (2007) to ensure analytical reliability. Using Nvivo 12.0 software, the interview transcripts were analyzed. Thematic analysis was employed to systematically examine and interpret the interview data. First, two researchers independently read the transcripts to familiarize themselves with the data. A two-stage coding procedure was then adopted.

Stage 1: Top-down coding. Agreed with the two researchers, Guided by Bandura's (1977) self-efficacy framework, student interview transcripts were initially code into four psychological processes: selection, cognitive, motivational, and emotional. For the teacher interview, two themes were established: academic achievement and learning engagement.

Stage 2: Bottom-up coding. The researchers conducted open coding to identify emergent ideas within each situation. The two researchers independently coded the first half of the transcripts, discussed discrepancies, refined the codebook, and reached consensus on the final coding scheme. The value of Cohen's Kappa coefficient for the analysis was 0.892, which met the acceptable standard. The remaining interview data were then coded using the finalized scheme.

Through this process, two themes related to academic achievement and learning engagement were generated from the teacher interview, and four themes corresponding to the self-efficacy processes were identified from the student interviews, providing insightful qualitative evidence to support and explain the quantitative findings.

## Results

### Impact of DI on academic achievement

The pre- and post-test academic performance scores of the two groups were analyzed using the Mann-Whitney U test. The pretest scores of the two groups of students showed no significant difference (see Table 6). After the experiment, significant differences were observed in the experimental group's post-test HOTS performance ( $Z=2.788$ ,  $p<0.01$ ) and total scores ( $Z=2.1665$ ,  $p<0.05$ ) compared to the control group. In the experimental group, the median of HOTS is 47.50, and the median of the total score is 85.00; in the control group, the median of HOTS is 36.25, and the median of the total score is 83.75, as detailed in Table 7. In addition, the median values suggest that participants in the experimental group (Mdn = 36.88) performed better on the level of LOTS than their counterparts in the control group (Mdn = 36.25). Nonetheless, the results also indicated that there was no statistical significance ( $Z=0.310$ ,  $p=0.757$ ) on the level of LOTS.

A paired-samples t-test was conducted for the pre- and post-tests of the experimental group (see Table 8). The t-statistic of the whole group was  $-46.60$ , with  $df = 63$  ( $p<0.001$ ). The effect size for the difference between the groups was calculated using Cohen's  $d$ , resulting in a value of 5.83, which is considered a high effect (Goulet-Pelletier & Cousineau, 2018). The t-statistics for the groups of students with third ( $M_{pre} = 12.44 \pm 6.07$ ,  $M_{post} = 76.37 \pm 11.46$ ), fourth ( $M_{pre} = 22.90 \pm 3.37$ ,  $M_{post} = 83.92 \pm 6.80$ ) and fifth grades ( $M_{pre} = 39.46 \pm 4.02$ ,  $M_{post} = 89.88 \pm 4.95$ ) were  $-30.62$ ,  $-33.60$ , and  $-33.17$ , respectively, and the p-values for each group were  $<0.00$ . These results signify a significant

improvement in academic achievements for students in the experimental group across the three grades. Based on Cohen's  $d$ , the effect sizes were large for all three grades, with the largest for the third grades.

The findings from the interviews provided more details that were able to corroborate the quantitative results. First, according to the teacher interview, the experimental group performed better than the control group in "completing prescribed learning tasks" and "actively inquiring higher-level tasks". Furthermore, the teacher noted that during class time, the experimental class "can basically complete the task", while in the control class, "at most 30% students can complete the task". In addition, students in the experimental group delved deeper, actively attempting to explain the phenomena and reflecting on the application of scientific principles. The teacher mentioned:

When inquiring the transfer of heat in metals, a child said, 'I think there must be a layer of insulation material on the outside of my insulated cup. Because when I pour hot water into this stainless steel insulated cup, I should have been able to feel the heat after a while if it weren't insulated, but now I haven't felt it.' This response demonstrates the student's ability to apply scientific reasoning to everyday objects and think critically about observed phenomena. In contrast, the students in the control group primarily described the phenomena mentioned in the activity sheet without further analysis, such as stating "Put the iron spoon in the hot soup, and the spoon also become[s] hot (interview, science teacher).

### Impact of DI on self-efficacy

In this study, the self-efficacy scores of the two groups were administered before and after the experiment were analyzed

**Table 6.** Mann-Whitney U test results for the group comparison of pretest.

	Experimental group (N=64)		Control group (N=65)		Mann-Whitney U	Z	p
	Mean ± SD	Mdn	Mean ± SD	Mdn			
LOTS	13.30 ± 8.06	12.50	12.69 ± 6.32	11.25	1910.00	-0.805	0.421
HOTS	11.50 ± 5.36	8.75	11.44 ± 5.78	12.50	2057.00	-0.108	0.914
Total	24.90 ± 12.02	23.13	24.13 ± 11.75	22.50	2067.50	-0.059	0.953

\* $p<0.05$ .

\*\* $p<0.01$ .

**Table 7.** Mann-Whitney U test results for the group comparison of post-test.

	Experimental group (N=64)		Control group (N=65)		Mann-Whitney U	Z	p
	Mean ± SD	Mdn	Mean ± SD	Mdn			
LOTS	36.62 ± 2.69	36.88	36.42 ± 3.15	36.25	2016.00	-0.310	0.757
HOTS	47.03 ± 7.75	47.50	37.67 ± 16.38	45.00	1490.50	-2.788	0.005**
Total	83.40 ± 9.77	85.00	74.09 ± 18.71	83.75	1622.00	-2.165	0.030*

\* $p<0.05$ .

\*\* $p<0.01$ .

**Table 8.** Paired-samples t-test results for pre and post-test in experimental group.

	Pretest		Post-test		df	t	p	Cohen's d
	M	SD	M	SD				
Fifth grade	39.46	4.02	89.88	4.95	20	-30.62	0.000**	6.68
Fourth grade	22.90	3.37	83.92	6.80	21	-33.60	0.000**	7.16
Third grade	12.44	6.07	76.37	11.46	20	-33.17	0.000**	7.24
Total	24.90	12.02	83.40	9.77	63	-46.60	0.000**	5.83

\* $p<0.05$ .

\*\* $p<0.01$ .

**Table 9.** Results of the self-efficacy questionnaire.

	Experimental group (N=64)		Control group (N=65)		Mann-Whitney U	Z	p
	Mean ± SD	Mdn	Mean ± SD	Mdn			
Before experiment	41.75 ± 1.88	42.00	41.77 ± 2.07	42.00	2059.50	-0.098	0.922
After experiment	42.95 ± 1.28	43.00	42.31 ± 3.66	44.00	1917.00	-0.783	0.434

\* $p < 0.05$ .\*\* $p < 0.01$ .**Table 10.** Results of the students' interview data.

	Experimental group		Control group	
	Aspect	Times	Aspect	Times
Selection process	Competitive tasks	11	Non-scored tasks	4
	More difficult	10	Less difficult	6
Cognitive process	Ability	5	Difficulty of the activity	6
	Effort	4	Lucky	4
	Length of participation in the inquiry	2	Class atmosphere	3
Motivational process	Interest	8	After-school rewards	20
	Desire to win	6	Peer dependence	2
Emotional process	Interesting	5	Boring	4
	Exciting	6	Nervous	6

using the Mann-Whitney U test. As shown in Table 9. The results of the pre-experiment self-efficacy questionnaire, revealed no significant difference among students in the two groups ( $Z=0.098$ ,  $p=0.922$ ). The median of the experimental group is 42.00, and the median of the control group is 42.00. Furthermore, the post-experiment self-efficacy questionnaire results for both groups still showed no significant difference ( $Z=0.783$ ,  $p=0.434$ ). The median of the experimental group is 43.00, and the median of the control group is 44.00.

The student interview data were analyzed based on Bandura's (1977) self-efficacy concepts. Bandura identified four major psychological processes of self-efficacy: selection, cognitive, motivational, and emotional. The interview results indicated that there was a significant difference in self-efficacy between the experimental group and the control group (see Table 10). First, regarding the selection process, students in the two groups exhibited differences in their preferences for learning task difficulty and organization. In the experimental group, students preferred competitive learning tasks, especially those that were more challenging. In contrast, the control group students favored non-scored tasks, i.e., those that were less challenging. Specifically, students in the experimental group showed a greater inclination to choose challenging tasks compared to the control group. Second, concerning cognitive processes, students in the two groups varied in their attributions of success or failure in the inquiry task. Students in the experimental group attributed their success or failure to internally controllable factors such as their own ability, effort, and length of participation in the inquiry. In contrast, students in the control group attributed externally uncontrollable factors such as the difficulty of the activity, luck, and class atmosphere. Third, regarding motivational processes, students in the experimental group were motivated to participate in the course by internal factors such as interest and the desire to win. In the control group, students were

more influenced by external motivations, including after-school rewards and peer dependence. Finally, in terms of emotional processes, students in the experimental group had more positive emotional feelings about the course, such as perceiving the course exploration as interesting or exciting. In contrast, the control group had relatively negative perceptions, stating that the course exploration bored them and made them nervous.

### Impact of DI on learning engagement

The data from the recorded in-class videos and teacher interviews were analyzed. The Mann-Whitney U test was used to analyze the ITIAS data for the two groups of teachers and students. Table 11 demonstrated that there were significant differences between the experimental and control groups in teacher talk, student talk, and silence ( $p < .05$ ). The experimental group showed lower proportions of teacher talk (Mdn = 0.17) and silence (Mdn = 0.06) compared with the control group (Mdn = 0.25 and 0.18, respectively), whereas student talk was substantially higher in the experimental group (Mdn = 0.64) than in the control group (Mdn = 0.45). Further analysis of teacher talk revealed no significant difference in indirect teacher talk, with similar medians for the experimental (Mdn = 0.11) and control groups (Mdn = 0.12). However, direct teacher talk differed significantly: the experimental group demonstrated a lower proportion (Mdn = 0.07) than the control group (Mdn = 0.13), indicating that teachers in the DI condition exerted less direct control over classroom discourse. Consistent with this shift, the proportion of student talk in the experimental group was significantly higher than that in the control group, reflecting greater student engagement under DI.

These findings were consistently reflected in the teacher's account in the interview. The teacher recounted that students in the experimental group were more "positive" and "actively engaged" in comparison to those in the control group. The teacher explained that she came with this conclusion based on her observation as she facilitated both groups. Specifically, with respect to the class that adopted DI, the teacher explained:

When I assigned learning tasks, the students in the experimental group quickly immersed themselves in the investigative activities and followed the learning task list. In the control group, however, there were instances where some student required more guidance from me before starting their tasks, and some students expressed, 'I think this is too easy; I want to do other (more challenging) activities'. Within such a classroom atmosphere, some students observed that others were uninterested, and they, too, lost motivation. I had to instruct the students separately and it was quite exhausting. (interview, science teacher).

**Table 11.** Results of the ITIAS encoding analysis descriptive data results.

Variables	Experimental group (N=8)		Control group (N=8)		Mann-Whitney U	Z	p
	Mean±SD	Mdn	Mean±SD	Mdn			
Teacher talk	0.16±0.05	0.17	0.25±0.07	0.25	11.00	-2.205	0.028*
Indirect impact	0.10±0.03	0.11	0.13±0.03	0.12	19.00	-1.365	0.195
Direct impact	0.07±0.02	0.07	0.12±0.04	0.13	6.00	-2.731	0.005**
Student talk	0.64±0.08	0.64	0.43±0.14	0.45	7.00	-2.626	0.007**
Silence	0.06±0.02	0.06	0.18±0.04	0.18	0.00	-3.361	0.000**
Technology	0.14±0.05	0.14	0.15±0.05	0.15	31.00	-0.105	0.959

\* $p < 0.05$ .\*\* $p < 0.01$ .

## Discussion

Our findings revealed that DI positively enhanced the experimental group students' learning engagement and academic achievements. Specifically, the students were highly engaged in the tailored and differentiated learning activities. As reported by the teacher, the students were both interested in and motivated to engage in the range of activities offered by the after-school science program, exhibiting the positive influence of DI on their overall learning engagement. Furthermore, this study demonstrates the benefits of DI in enhancing the learning performance of students in the cross-grade levels classroom. Although the self-efficacy post-test scores did not reveal significant differences, insights obtained from student interviews presented a contrasting picture.

With respect to RQ1, our study found that students in the DI group performed higher on HOTS compared to the control group, which is consistent with previous study in traditional primary and secondary school classrooms, showing an association between DI and higher order learning achievements (Chen & Shih, 2025; Salvador, 2023). Furthermore, this study extends prior studies by showing higher HOTS scores in the DI group in a cross-grade after-school science program. DI can assist teachers in designing learning tasks that align with students' Zone of Proximal Development (ZPD), thereby ensuring that instructional activities are appropriately challenging and developmentally responsive (Vygotsky, 1980). By offering tasks just beyond students' current capabilities—accompanied by sufficient scaffolding—DI helps learners engage in productive struggle while accumulating successful learning experiences. Additionally, although previous research has shown that the use of DI in heterogeneous benefits lower-achieving students the most (Kotob & Ali Abadi, 2019), the heterogeneous classrooms in this study included students from different grades. The results indicated that students at a starting of younger-grade (third grade in this study), but not necessarily the lower-achieving, benefited the most from DI compared with other students in higher grades. Due to their relatively lower cognitive and self-regulation abilities, younger-grade students rely more on the guidance and support of teachers and peers (Jiao, 2016). DI, through strategies such as group collaboration, tailored content, process, product, and learning environment, provides essential social scaffolding that significantly enhances knowledge construction for them. Furthermore, considering that students in heterogeneous classrooms particularly benefit from enhanced cognitive

activation and a supportive learning environment (Decristan et al., 2017). In this study, the implementation of DI successfully provided both of these elements, engaging students in cognitively challenging tasks while offering scaffolded support and collaborative opportunities. These insights suggest that the application of DI could be further extended to diverse and heterogeneous learning contexts, including after-school cross-grade programs, to maximize academic achievement and foster higher-order thinking skills among students.

In relation to RQ2, although the self-efficacy post-test scores did not reveal significant differences between the experimental and control groups, insights obtained from student interviews presented a contrasting picture, suggesting that participants in the experimental group appeared to experience greater self-efficacy in terms of selection, cognitive, motivational, and emotional dimensions. This may be influenced by factors such as the duration of after-school program and the sensitivity of measurement tools. Additionally, Carroll et al. (2024) found that primary school students tend to report higher levels of science learning self-efficacy on questionnaires than they actually demonstrate during scientific inquiry. This tendency toward overestimation may help explain why both the experimental and control groups in our study showed relatively high levels of self-efficacy, and why the instrument may not have been sensitive enough to detect differences between the two groups. According to Bandura's (1993) theory, by designing content, process, product, and learning environment align with students' ZPD, DI providing repeated opportunities for successful experiences to students, which plays an important role to students' self-efficacy. Furthermore, social comparison theory (Festinger, 1954) posits that students develop self-evaluations by observing the performance of peers who are similar to themselves. In the experimental group, the weekly adjustment of group placements, together with the designation of peer tutors, provided students with powerful and attainable role models. These experiences offered clear and observable pathways to success, reinforced students' perceptions of their own capabilities, which may be associated with the development of higher self-efficacy.

With regard to RQ3, the results of the class video analysis and teacher interviews further highlighted that DI was associated with more active student engagement. Students in the experimental group talked more and were more positive and actively engaged than those in the control group. These findings align with prior studies showing that DI has been associated with greater engagement by integrating choice,

flexibility, and relevance into students' learning (Ramaila, 2025). At the same time, our finding contrasts with the finding of Zens (2021), who reported that DI had little impact on student engagement in ninth-grade English language arts classrooms. This difference can be interpreted through the lens of self-determination theory (Deci & Ryan, 1985), which posits that intrinsic motivation arises from the fulfillment of three basic psychological needs: autonomy, competence, and relatedness. In the context of this study, compared with compulsory in-class English language arts learning, the optional nature of the after-school program satisfied students' need for autonomy. Building on this foundation, DI addressed the needs for competence and relatedness by offering tailored science activities and tasks alongside a supportive and positive learning environment. Together, these elements may have contributed to students' intrinsic motivation, which may help explain the higher levels of engagement observed.

### Conclusions & limitations

This study examined the use of DI on student learning within a cross-grade after-school science program at a primary school in northwest China. While existing research has widely supported the positive associations reported for DI, few studies have explored its implementation in cross-grade after-school science contexts. The results indicated that DI was associated with students' academic achievement—particularly among younger-grade students, and showed greater learning engagement. Notably, DI was found associated with higher-order cognitive outcomes, aligning with the goals of deeper science learning. Although the self-efficacy questionnaire data did not show statistically significant differences on students' self-efficacy, qualitative insights gathered through student interviews revealed a different pattern, with many students expressing positive attitude.

This study employed a mixed-methods approach, collecting four types of data encompassing both quantitative and qualitative dimensions. However, the study's findings should be interpreted with caution due to certain limitations—most notably, the relatively small sample size and the fact that all participants were drawn from a single school, which constrains the representativeness of the sample and limits the external validity of the conclusions. Future research should consider conducting large-scale studies across multiple regions and school types to test the generalizability of these findings and to further explore the potential moderating effects of contextual factors such as regional differences, school characteristics, and cultural background. Moreover, the reliance on a single teacher to deliver interventions in both groups introduces the risk of the experimenter effect. Although rigorous instructional fidelity checks were performed, the teacher's awareness of the research objectives might have subtly influenced the implementation. Future research should involve a more diverse cohort of teachers across different school settings to enhance ecological validity and mitigate potential bias associated with individual teacher characteristics. In addition, this study explored the use of

data from intelligent learning platforms to support teachers in implementing flexible grouping strategies, with the aim of developing instructional approaches better aligned with individual learner differences and advancing the application of DI. However, due to practical considerations related to classroom management and instructional efficiency, the grouping approach primarily relied on homogeneous groupings based on students' learning performance. While this approach facilitated targeted instruction, it may have limited the potential benefits of heterogeneous grouping—such as cognitive stimulation and peer scaffolding among students with diverse learning styles and abilities. Future research should investigate the comparative effects of homogeneous versus heterogeneous grouping on students' learning performance, emotional motivation, and quality of collaboration, in order to refine group design and optimize its role in supporting effective differentiated teaching. Furthermore, this study find that even younger-grade students can benefit meaningfully from DI, an outcome that offers valuable pedagogical insights. However, the relatively short duration of the study and the absence of longitudinal tracking limit the ability to evaluate the sustained effects of DI on students' long-term development, particularly in areas such as cognitive strategy use, self-regulation, and academic resilience. To gain a more comprehensive understanding of DI's developmental impact, future research is encouraged to adopt longitudinal designs that systematically track student growth in cross-grade DI environments. Such studies could provide deeper insight into the educational potential of DI and inform more effective, developmentally responsive instructional practices.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### Funding

This work was supported by the National Social Science Foundation of China Youth Program in Education [Grant number: CCA240257].

### ORCID

Shu Zhao  <http://orcid.org/0000-0003-2349-6830>  
 Wei Liu  <http://orcid.org/0009-0007-8198-3537>  
 Yuxin Zuo  <http://orcid.org/0009-0009-6353-6367>  
 Grace Yue Qi  <http://orcid.org/0000-0002-5757-4203>  
 Wanshan Hu  <http://orcid.org/0000-0003-2281-7006>

### Data availability statement

The data that support the findings of this study are openly available in Harvard Dataverse at <https://doi.org/10.7910/DVN/GE2SG6>, reference number 10.7910/DVN/GE2SG6.

### References

- Allen, P. J., Chang, R., Gorrall, B. K., Waggenspack, L., Fukuda, E., Little, T. D., & Noam, G. G. (2019). From quality to outcomes: A national study of afterschool STEM programming. *International*

- Journal of STEM Education*, 6(1), 37. <https://doi.org/10.1186/s40594-019-0191-2>
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al. (2000). A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives, abridged edition. *European Legacy*, 114(458), 1013–1014. <https://doi.org/10.1093/ehr/114.458.1013>
- Ariss, L. D. (2017). *Differentiated instruction: An exploratory study in a secondary mathematics classroom* [Doctoral dissertation, University of Toledo].
- Banchi, H., & Bell, R. (2008). The many levels of inquiry. *Science and Children*, 46(2), 26–29.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117–148. [https://doi.org/10.1207/s15326985ep2802\\_3](https://doi.org/10.1207/s15326985ep2802_3)
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295x.84.2.191>
- Boggs, J. K. (2024). *The effect of an after-school outdoor adventure education program on high school students outdoor recreation perspective* (Order No. 31330400) [Doctoral dissertation, The University of North Carolina at Greensboro]. ProQuest Dissertations & Theses Global. 3095285956. <https://www.proquest.com/dissertations-theses/effect-after-school-outdoor-adventure-education/docview/3095285956/se-2>
- Bloom, B. S. (1956). *Taxonomy of educational objectives, Handbook I: The cognitive domain*. David McKay Co Inc.
- Carroll, S., McCauley, V., & Grenon, M. (2024). Science self-efficacy beliefs of upper primary students in Ireland. *International Journal of Science Education*, 46(6), 503–523. <https://doi.org/10.1080/09500693.2023.2245947>
- Chen, C.-H., & Shih, D.-J. (2025). Facilitating EFL students' class engagement, motivation, self-efficacy, and achievements: Adopting differentiated instruction in self-regulated flipped learning. *Educational Technology Research and Development*, 73(5), 3201–3221. <https://doi.org/10.1007/s11423-025-10520-y>
- Creswell, J. W., & Clark, V. L. P. (2007). *Designing and conducting mixed methods research*. Sage publications. <https://doi.org/10.1111/j.1753-6405.2007.00096.x>
- Dack, H., & Tomlinson, C. (2015). Inviting all students to learn. *Educational Leadership: Journal of the Department of Supervision and Curriculum Development, N.E.A.*, 72, 10–15.
- Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, 19(2), 109–134. [https://doi.org/10.1016/0092-6566\(85\)90023-6](https://doi.org/10.1016/0092-6566(85)90023-6)
- Decristan, J., Fauth, B., Kunter, M., Büttner, G., & Klieme, E. (2017). The interplay between class heterogeneity and teaching quality in primary school. *International Journal of Educational Research*, 86, 109–121. <https://doi.org/10.1016/j.ijer.2017.09.004>
- Durlak, J., & Weissberg, R. (2007). *The impact of after-school programs that promote personal and social skills*. Collaborative for Academic, Social, and Emotional Learning.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117–140. <https://doi.org/10.1177/001872675400700202>
- Flanders, N. A. (1970). *Analyzing teacher behavior*. Addison-Wesley P. C.
- Goodnough, K. (2010). Investigating pre-service science teachers' developing professional knowledge through the lens of differentiated instruction. *Research in Science Education*, 40(2), 239–265. <https://doi.org/10.1007/s11165-009-9120-6>
- Goulet-Pelletier, J.-C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: The Cohen'sd family. *The Quantitative Methods for Psychology*, 14(4), 242–265. <https://doi.org/10.20982/tqmp.14.4.p242>
- Greco, Y. V. (2023). DI: Curriculum and resources provide a roadmap to help English teachers meet students' needs. *Teaching and Teacher Education*, 125, 104064. <https://doi.org/10.1016/j.tate.2023.104064>
- Gu, X. Q., & Wang, W. (2004). 支持教师专业发展的课堂分析技术新探索 [New exploration of classroom analysis technology to support teachers' professional development]. *中国电化教育 [China Educational Technology]*, 7, 4.
- Haelermans, C., Ghysels, J., & Prince, F. (2015). Increasing performance by differentiated teaching? experimental evidence of the student benefits of digital differentiation. *British Journal of Educational Technology*, 46(6), 1161–1174. <https://doi.org/10.1111/bjet.12209>
- Hattie, J. (2008). *Visible learning*. Routledge eBooks. <https://doi.org/10.4324/9780203887332>
- Hayes, K. (2018). *Exploration of a collaborative non-formal science education program in Australian secondary schools* [Doctoral dissertation, Australian National University]. <https://hdl.handle.net/1885/144664>
- Horak, V. M. (1981). A meta-analysis of research findings on individualized instruction in mathematics. *The Journal of Educational Research*, 74(4), 249–253. <https://doi.org/10.1080/00220671.1981.10885318>
- Jiao, X. (2016). The development of children's self-regulation in elementary school. *Advances in Psychology*, 06(04), 398–404. <https://doi.org/10.12677/ap.2016.64052>
- Karlsudd, P. I. (2023). Inclusion through participation: Fostering pupils' feelings of belonging in swedish after-school care. *Education Sciences*, 13(4), 376. <https://doi.org/10.3390/educsci13040376>
- Kolb, D. (1984). Experiential learning: Experience as the source of learning and development. *Journal of Business Ethics*, 1, 20–38.
- Kotob, M. M., & Ali Abadi, M. (2019). The influence of differentiated instruction on academic achievement of students in mixed ability classrooms. *International Linguistics Research*, 2(2), p8. <https://doi.org/10.30560/ilr.v2n2p8>
- Lai, C.-P., Zhang, W., & Chang, Y.-L. (2020). Differentiated Instruction enhances sixth-grade students' mathematics self-efficacy, learning motives, and problem-solving skills. *Social Behavior and Personality: An International Journal*, 48(6), 1–13. <https://doi.org/10.2224/sbp.9094>
- Liu, X., & Yuan, X. (2024). Analysis of distortions in the implementation of the "double reduction" policy. *Journal of Education and Educational Research*, 8(3), 57–60. <https://doi.org/10.54097/m3rbdq10>
- Maeng, J. L. (2017). Using technology to facilitate differentiated high school science instruction. *Research in Science Education*, 47(5), 1075–1099. <https://doi.org/10.1007/s11165-016-9546-6>
- Magaji, A., Ade-Ojo, G., & Bijlhout, D. (2022). The impact of after school science club on the learning progress and attainment of students. *International Journal of Instruction*, 15(3), 171–190. <https://doi.org/10.29333/iji.2022.15310a>
- McMahon, K., Berger, M., Khalsa, K. K., Harden, E., & Khalsa, S. B. S. (2021). A non-randomized trial of kundalini yoga for emotion regulation within an after-school program for adolescents. *Journal of Child and Family Studies*, 30(3), 711–722. <https://doi.org/10.1007/s10826-021-01911-9>
- Moreno, N. P., Tharp, B. Z., Vogt, G., Newell, A. D., & Burnett, C. A. (2016). Preparing students for middle school through after-school STEM activities. *Journal of Science Education and Technology*, 25(6), 889–897. <https://doi.org/10.1007/s10956-016-9643-3>
- Newell, A. D., Zientek, L. R., Tharp, B. Z., Vogt, G. L., & Moreno, N. P. (2015). Students' attitudes toward science as predictors of gains on student content knowledge: Benefits of an after-school program. *School Science and Mathematics*, 115(5), 216–225. <https://doi.org/10.1111/ssm.12125>
- OECD. (2023, May 1). *PISA 2025 science framework*. [https://pisa-framework.oecd.org/science-2025/assets/docs/PISA\\_2025\\_Science\\_Framework.pdf](https://pisa-framework.oecd.org/science-2025/assets/docs/PISA_2025_Science_Framework.pdf)
- Pablico, J. R., Diack, M., & Lawson, A. (2017). Differentiated instruction in the high school science classroom: Qualitative and quantitative analyses. *International Journal of Learning, Teaching and Educational Research*, 16(7), 30–54.
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40. <https://doi.org/10.1037/0022-0663.82.1.33>
- Pozas, M., Letzel, V., & Schneider, C. (2020). Teachers and differentiated instruction: Exploring differentiation practices to address student

- diversity. *Journal of Research in Special Educational Needs*, 20(3), 217–230. <https://doi.org/10.1111/1471-3802.12481>
- Ramaila, S. (2025). Unveiling the potential: A systematic review on harnessing the affordances of differentiated instruction. *Journal of Teaching and Learning*, 19(2), 41–56. <https://doi.org/10.22329/jtl.v19i2.8561>
- Sahin, A., Ayar, M. C., & Adiguzel, T. (2014). STEM related after-school program activities and associated outcomes on student learning. *Educational Sciences: Theory & Practice*, 14(1), 309–322. <https://doi.org/10.12738/estp.2014.1.1876>
- Salvador, N. T. (2023). Differentiated instruction and higher order thinking skills of grade 12 students in understanding culture. *International Journal of Advanced Multidisciplinary Studies*, 3(6), 425–451.
- Shea, M. V., Jurow, A. S., Schiffer, J., Escudé, M., & Torres, A. (2023). Infrastructural injustices in community-driven afterschool STEAM. *Journal of Research in Science Teaching*, 60(8), 1853–1878. <https://doi.org/10.1002/tea.21852>
- Smale-Jacobse, A. E., Meijer, A., Helms-Lorenz, M., & Maulana, R. (2019). Differentiated Instruction in secondary education: A systematic review of research evidence. *Frontiers in Psychology*, 10, 2366. <https://doi.org/10.3389/fpsyg.2019.02366>
- Subban, P. (2006). Differentiated instruction: A research basis. *International Education Journal*, 7, 935–947. <https://files.eric.ed.gov/fulltext/EJ854351.pdf>
- Suprayogi, M., & Valcke, M. (2016). Differentiated instruction in primary schools: Implementation and challenges in Indonesia. *PONTE International Scientific Researches Journal*, 72(6), 2–18. <https://doi.org/10.21506/j.ponte.2016.6.1>
- Suwastini, N. K. A., Rinawati, N. K. A., Jayantini, I. G. A. S. R., & Dantes, G. R. (2021). Differentiated Instruction across EFL classrooms: A conceptual review. *Tell-Us Journal*, 7(1), 14–41. <https://doi.org/10.22202/tus.2021.v7i1.4719>
- Tomlinson, C. A. (2000). *Differentiation of instruction in the elementary grades*. ERIC Digest.
- Tomlinson, C. A., & Moon, T. R. (2013). *Assessment and student success in a differentiated classroom*. ASCD.
- Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners*. Ascd.
- Tomlinson, C. A., & McTighe, J. (2006). *Integrating differentiated instruction and understanding by design: Connecting content and kids*. Association for Supervision and Curriculum Development.
- Vandell, D. L., Reisner, E. R., & Pierce, K. M. (2007). *Outcomes linked to high-quality afterschool programs: Longitudinal findings from the study of promising afterschool programs*. Policy Studies Associates, Inc.
- van Geel, M., Keuning, T., Frèrejean, J., Dolmans, D., van Merriënboer, J., & Visscher, A. J. (2019). Capturing the complexity of differentiated instruction. *School Effectiveness and School Improvement*, 30(1), 51–67. <https://doi.org/10.1080/09243453.2018.1539013>
- Veldman, S. L. C., Jones, R. A., Stanley, R. M., Cliff, D. P., Vella, S. A., Howard, S. J., Parrish, A.-M., & Okely, A. D. (2020). Promoting physical activity and executive functions among children: A cluster randomized controlled trial of an after-school program in Australia. *Journal of Physical Activity & Health*, 17(10), 940–946. <https://doi.org/10.1123/jpah.2019-0381>
- Vygotsky, L. S. (1980). *Mind in society*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Zens, A. (2021). *The impact of differentiated learning activities on student engagement and motivation in the English language arts classroom* (Publication No. 611) [Doctoral dissertation]. Minnesota State University]. Dissertations, Theses, and Projects. <https://red.mnstate.edu/thesis/611>