ISSN 2618-1576 Vol. 7, Issue 5: 95-101, DOI: 10.25236/FSR.2025.070515

Empirical Study on the Teaching Effectiveness of Physical Education Classroom Driven by DeepSeek and Multi Generative Intelligent Tools

Luojun Li*

Foshan Sanshui Xinan Sub-district Jinben Middle School, Foshan 528100, Guangdong, China
**Corresponding author*

Abstract: *This study aims to systematically explore the empirical impact of a physical education teaching model based on the deep integration of DeepSeek big language model and multi-dimensional generative intelligent tools (such as image and video generation tools) on the teaching effectiveness of middle school students. The study adopted the "non equivalent control group pre-test design" in quasi experimental research, selecting two parallel classes in the second year of junior high school at Jinben Middle School, divided into an experimental class (n=45) and a control class (n=45), for a one semester (16 weeks) teaching intervention. The experimental class implements the "DeepSeek+" teaching mode with "personalized pre class preparation, human-machine collaboration during class, and precise extension after class" as the core, while the control class maintains the traditional teaching mode. This paper collect data through various tools such as standardized sports skill tests, sports theory knowledge papers, sports learning interest scales, deep learning process questionnaires, and semi-structured interviews. The research results showed that after controlling for pre-test differences, the experimental class students had significantly higher post test scores in multiple dimensions, including motor skill mastery level ($p<0.01$), sports theory score ($p<0.001$), sports learning interest ($p<0.01$), and deep learning ability (such as critical thinking, communication and collaboration, $p<0.05$), compared to the control class. Qualitative interview data further revealed that this model effectively enhances students' self-efficacy and intrinsic motivation by providing immediate and accurate feedback, implementing personalized learning paths, and creating high immersion learning contexts. The "DeepSeek+" physical education teaching model can comprehensively and multi-dimensional improve teaching effectiveness, not only in terms of knowledge and skills, but also in terms of stimulating interest and promoting the development of higher-order thinking abilities, showing great potential. This study provides solid empirical evidence for the deep integration of generative artificial intelligence and physical education, and puts forward practical suggestions for the construction of future smart physical education classrooms.*

Keywords: *Deep Seek; Generative AI Physical Education Teaching; Teaching Effectiveness*

1. Introduction

The physical education and health curriculum is a key link in comprehensively enhancing the core literacy of young people and laying the foundation for lifelong health. However, the traditional physical education classroom teaching model has exposed many irreconcilable contradictions and limitations in long-term practice[1]. Firstly, there is a huge tension between the standardization of "teaching" and the individualization of "learning". A physical education teacher faces dozens of students, making it difficult to observe and provide personalized guidance on each student's motor skills, resulting in extensive teaching guidance and the coexistence of "not having enough to eat" and "not keeping up" phenomena. Secondly, there is a serious disconnect between theoretical teaching and practical operation. Theoretical knowledge about sports physiology, mechanics principles, nutrition and health is often presented in dry oral or written form, which is difficult to resonate with students and leads to students' "knowing what they are but not knowing why they are". Again, the teaching evaluation method is subjective and relies heavily on teachers' summative and experiential evaluations, lacking process and data-driven support, making it difficult to truly and comprehensively reflect students' growth and efforts. In recent years, the explosive development of large-scale language models such as DeepSeek and ChatGPT, as well as image and video generative artificial intelligence technologies such as DALL-E, Midjourney, and Sora, has brought historic opportunities to solve the above-mentioned difficulties[2]. These technologies demonstrate powerful abilities in natural language understanding and generation, knowledge integration and reasoning, and multimodal content creation, making it possible to build a highly personalized, interactive, and intelligent learning environment [3]. Currently, the academic community has reached a consensus on the trend of AI empowering education, and a large number of theoretical discussions and conceptual framework designs have emerged [4]. However, focusing on the specific field of physical education, especially the effectiveness of the comprehensive teaching model centered on the big language model and integrating multiple generative tools, most existing research still remains at the level of logical deduction, prospect outlook, and preliminary case analysis. A core and unresolved question is whether this deeply integrated "DeepSeek+" teaching model can produce statistically significant and multidimensional improvements in teaching effectiveness compared to traditional models? What is the magnitude of its effect? What is its internal mechanism of action?

This study aims to conduct a systematic and in-depth empirical examination of the above issues through a rigorous quasi experimental design. The specific research objectives include: quantitatively comparing the differences between the "DeepSeek+" model and the traditional model in core outcome variables such as students' motor skills, sports theory knowledge, and learning interests. This paper further explore the potential promoting effect of this model on students' deep learning abilities, such as problem-solving, critical thinking, and collaborative skills [5]. By using qualitative research methods, this study aims to deeply explore students' learning experiences and perceptions under this mode, and reveal the underlying mechanisms that enhance teaching effectiveness. The theoretical significance of this study lies in pushing the research on the application of generative AI education from macro discourse to micro empirical research, providing solid data support from the teaching front line for the "AI+sports" education theory, and enriching the interdisciplinary research between educational technology and sports teaching theory. Its practical significance lies in providing a set of operable and assessable teaching reform plans and decision-making basis for frontline physical education teachers and educational managers, promoting the digital transformation and intelligent upgrading of physical education classrooms.

2. Literature review and research framework

2.1 Evolution of the application of generative intelligence tools in education

The educational application of generative AI is moving from "tool assisted" to "ecological reshaping". Early applications were mostly isolated tools, such as for grading essays or generating simple questions. With the emergence of big language models, their roles have evolved into "conversational mentors" and "content generators". Currently, the forefront of research emphasizes the collaborative integration of multiple tools. For example, we use a big language model for teaching logic design and personalized dialogue, while calling image and video generation tools to visualize knowledge presentation, combined with code generation tools for data analysis, to form a complete teaching solution. The "DeepSeek+" model in this study is the concrete practice of this cutting-edge trend in the field of sports.

2.2 Multidimensional evaluation system for physical education teaching effectiveness

The traditional evaluation of physical education teaching effectiveness focuses too much on achieving skill standards, which is one-sided. Modern educational evaluation theory emphasizes comprehensive evaluation [6]. This study draws on Kirk et al.'s classification of physical education objectives and constructs a multidimensional evaluation framework: (1) in the field of motor skills, including the degree of automation of basic motor skills and mastery of complex skills; (2) Cognitive domain: including understanding, application, and analysis of knowledge related to sports; (3) Emotional and social domains: including learning interests, motivation, attitudes, and collaborative communication skills. This framework ensures that research can comprehensively and stereoscopically capture the true effects of teaching modes.

2.3 Theoretical research gaps and the positioning of this study

Although research has confirmed the positive effects of video feedback, wearable devices, and other technologies in physical education teaching, there is still a lack of comprehensive intervention research based on large language models as the cognitive core[7]. The value of the big language model lies not only in content generation, but also in its ability as a "cognitive partner" to engage in deep dialogue with students and guide them to engage in metacognitive reflection, which traditional technological tools do not possess. Therefore, this study combines DeepSeek's conversational reasoning ability with the

ISSN 2618-1576 Vol. 7, Issue 5: 95-101, DOI: 10.25236/FSR.2025.070515

perceptual generation ability of multiple tools, aiming to fill this theoretical research gap. The conceptual framework is shown in the following figure: the framework is driven by "DeepSeek+Multiple Tools" and reshapes the "pre class in class post class" teaching process, ultimately affecting the multi-dimensional teaching effect of "skills cognition emotion ability".

3. Research design and methods

3.1 Research paradigm and experimental design

This study adopts a quasi experimental research paradigm and utilizes a "non equivalent control group pre-test and post test design". The reason for choosing this paradigm is that in a real school environment, it is not possible to completely randomly allocate classes, and this design can effectively control most of the internal validity threats.

3.2 Research object and sampling

The research was conducted in a provincial demonstration high school in a certain city. This paper select the two classes in the second year of junior high school that are closest in demographic variables (gender, age) and total physical fitness test scores from the previous semester among all parallel classes: Class 3 and Class 4. Using the coin toss method, randomly select (3) class as the experimental group (n=45) and (4) class as the control group (n=45). Before the experiment, conduct baseline tests on all students to ensure that there are no significant differences between the two groups of students in various pre-test indicators. All participating students and parents have signed informed consent forms.

3.3 Experimental intervention plan

The experimental period is the first semester of the 2024-2025 academic year, with a total of 16 weeks and 3 physical education classes per week. Control class: it implement traditional physical education teaching mode. The process is: teacher's unified explanation and demonstration (about 10 minutes) ->student collective imitation practice ->student group practice, teacher's tour guidance ->collective summary before the end of the class. Theoretical teaching mainly relies on teachers' oral narration. Experimental class: it implement the "DeepSeek+" physical education teaching model, with the following core components:

Before class: The intelligent preview and learning situation diagnosis teacher uses DeepSeek to input teaching objectives and student level parameters, generating a hierarchical preview material package. For example, this study focuses on the theme of "basketball jump shot" and generates graphic guides and slow videos for beginners, with a particular emphasis on hand shape and standing position; It generates the principle analysis of whole-body coordination and parabolic optimization for advanced learners. Students preview through the smart sports learning platform and submit their questions online. DeepSeek conducts cluster analysis on the questions of the whole class and generates a learning report for teachers to refer to in lesson preparation.

In class: Human computer collaboration and precise feedback theory section: Teachers no longer simply teach, but use AI generated 3D muscle working principle animations, tactical running simulation videos, and other situational teaching methods. At the same time, an "AI Q&A corner" is set up in the classroom, where students can have conversations with DeepSeek about difficult questions at any time. Practice session: Students use tablets equipped with AI visual analysis applications to record their own action videos (such as standing long jump, soccer shot) during practice. The application is based on open-source pose estimation libraries such as OpenPose and MediaPipe, which generate bone points in real-time and compare them with standard action models, providing real-time visual and data-driven feedback (such as "jump angle too small by 5 °" and "insufficient swing amplitude"). For the professional terminology or improvement suggestions in the feedback report, students can request DeepSeek to provide oral explanations or targeted auxiliary exercises.

After class: The personalized extension and metacognitive cultivation learning platform is based on students' performance data in class (such as AI analysis reports and practice frequency), and generates and pushes "one person, one case" post class tasks through DeepSeek. This includes not only physical skill exercises (such as "3 home exercises for your ankle weakness"), but also metacognitive guidance questions (such as "Looking back at your learning today, which aspect do you think you have made the greatest progress in? Why?"), promoting student reflection.

ISSN 2618-1576 Vol. 7, Issue 5: 95-101, DOI: 10.25236/FSR.2025.070515

3.4 Measurement tools and reliability and validity

Sports skills test: Select the "National Student Physical Health Standards" and the school's regular skill assessment items, "Basketball half court back and forth dribbling layup (timing)" and "Volleyball cushion up to standard (counting)", as the evaluation content. A scoring group consisting of three senior physical education teachers who are not familiar with the grouping situation will use a unified scoring standard for evaluation. This paper calculate the intra group correlation coefficient (ICC) to evaluate the reliability of the raters. The ICC values for both items are greater than 0.85, indicating good reliability. Sports Theory Knowledge Test Paper: Self designed "Sports and Health Theory Knowledge Test Paper", covering the principles of sports skills, competition rules, sports nutrition, injury prevention, and other topics learned this semester. The test paper has been reviewed by three subject experts and has good content validity. Reliability analysis was conducted on the pre-test data, and the Cronbach's alpha coefficient was 0.81, indicating acceptable reliability. Sports Learning Interest Scale: Adopting the widely used Chinese revised version of the "Middle School Students' Sports Learning Interest Level Scale" both domestically and internationally, it includes four dimensions: positive emotions, learning engagement, independent exploration, and attention to sports. The overall Cronbach's alpha coefficient of this scale in the sample of this study is 0.89, and the coefficients of each dimension are all above 0.75, indicating ideal reliability. Deep Learning Process Questionnaire: Using the "Deep Learning Process Questionnaire" developed by Nelson Laird et al. and revised in Chinese, it mainly measures dimensions such as critical thinking, integrated learning, and reflective learning. The Cronbach's alpha coefficient of this scale in this study was 0.83. Semi structured interview outline: A self-designed interview outline is used to conduct sample interviews with experimental class students after the experiment, in order to gain a deeper understanding of their acceptance of AI tools, changes in learning experience, gains and challenges in self-awareness, etc.

3.5 Data analysis methods

This article uses SPSS 26.0 to process quantitative data. Firstly, it conduct independent sample t-test on the pre-test data of the two groups to ensure homogeneity between the groups. Secondly, in order to accurately evaluate the effectiveness of the teaching mode and eliminate the influence of pre-test scores on post test scores, covariance analysis (ANCOVA) was conducted on various post test scores, with pre-test scores as covariates. For qualitative data such as interviews, the thematic analysis method is used for encoding, summarizing, and analyzing to supplement and explain quantitative results.

4. Results

4.1 Homogeneity test results

Before the experimental intervention, pre tests were conducted on the motor skills (basketball, volleyball), theoretical grades, learning interests, and deep learning abilities of students in two classes. The independent sample t-test results showed that there were no statistically significant differences between the two groups in all items ($p>0.05$), confirming the comparability between the experimental class and the control class at the starting point of the experiment.

4.2 Quantitative results

The comparison of inter group differences in post test scores was conducted using pre test scores as covariates, and covariance analysis was performed on post test scores. The results are shown in Table 1.

As can be seen from the Table 1:

Motor Skills: In basketball (speed) and volleyball (number of passes), the post-test scores of the experimental class were significantly higher than those of the control class ($F=16.83$, $p<0.01$; $F=14.25$, $p<0.01$), with effect sizes (partial η^2) reaching 0.161 and 0.141, respectively, indicating large effect sizes. This suggests that the difference in teaching models was the primary cause of the skill disparities.

Theoretical score: The experimental class showed the most significant improvement in theoretical score ($F=25.67$, $p<0.001$), with a high effect size of 0.231, indicating that the "DeepSeek+" model has outstanding advantages in promoting knowledge understanding and mastery.

Learning interest: The improvement of learning interest among students in the experimental class was significantly higher than that in the control class (F=8.92, p<0.01), with an effect size of 0.093, close to the moderate equivalent stress.

Deep learning ability: In terms of questionnaire scores during the deep learning process, the experimental group also scored significantly higher than the control group (F=5.14, p<0.05). Although the effect size (0.057) was moderate to small, this result was statistically significant, indicating that this model has a positive effect on promoting students' higher-order thinking development.

Table 1 Covariance analysis (ANCOVA) results of post test scores between experimental and control groups (M ± SD)

Dependent variable	Group	Before measurement (M±SD)	Aftertest (M±SD)	F price	P price	Inclined to one side η²
Basketball skills (s)	Experimental group	22.45±3.12	19.76±2.45	16.83	0.000**	0.161
	Control group	22.81±2.87	21.63±2.76			
Volleyball skills (points)	Experimental group	25.58±6.21	35.82±5.07	14.25	0.000**	0.141
	Control group	24.91±5.83	28.35±7.95			
Theoretical score (points)	Experimental group	72.31±8.45	85.64±6.18	25.67	0.000**	0.231
	Control group	71.82±9.11	76.35±7.95			
Learning interest (points)	Experimental group	68.47±7.76	80.18±6.48	8.92	0.004**	0.093
	Control group	67.89±8.23	74.79±7.87			
Deep learning (split)	Experimental group	70.12±6.54	76.89±5.23	5.14	0.026*	0.057
	Control group	69.78±7.01	73.45±6.88			

Note: p < 0.05, *p < 0.01; partial η² is the effect size indicator, with 0.01 indicating a small effect, 0.06 a medium effect, and 0.14 a large effect.

4.3 Qualitative results

Theme Analysis of Student Experience: The interview data of 6 experimental class students were sorted and analyzed, and three core themes were summarized:

Topic 1: "From Ambiguity to Clarity" - Precision and Cognitive Transparency of Feedback. Students generally emphasize the clarity brought by AI feedback. (Middle aged student) said, "In the past, my teacher told me that my movements were stiff, and I didn't know exactly where they were stiff. Now, the tablet will mark my joint angles, tell me that my knee arthritis is not enough, and show me a standard animation comparison. It will be clear at a glance." This data-driven and visual feedback makes the originally implicit and vague feeling of movements become explicit and concrete, greatly reducing cognitive load.

Topic 2: "My Practice is My Own" - Personalization and Autonomy in Learning. Several students mentioned the positive impact of personalized programs. (Hou Jinsheng) said, "The homework given to me by AI is ball cushioning and wrist strength exercises, because the report says I am weak in these areas. It feels like helping me solve problems alone, giving me more motivation to complete them. ”(Excellent student) said, "I can see my small progress data every time, and I can challenge DeepSeek for deeper problems, such as tactical choices under different defenses, which makes me feel that there is no ceiling to my learning. ”

Theme 3: "Teaching Like Playing a Game" - The Intrinsic Motivation and Immersion in the Context Students liken this learning experience to a "level game". Mentioning: "Every practice is like 'brushing copies', you can immediately see' scores' and 'evaluations', just wanting to do better each time." Instant feedback, clear goals, and personalized challenges together create a highly immersive learning environment, transforming external demands into internal exploration desires.

5. Discussion

5.1 Comprehensive discussion on research findings

The results of this study comprehensively and strongly confirm the superiority of the "DeepSeek+" physical education teaching model. Its success can be attributed to the following core mechanisms: firstly, it has achieved "personalized education on a large scale". The problem of "teaching according to students' aptitude" that traditional teaching cannot solve has been effectively solved through AI technology. The combination of DeepSeek and visual analysis tools is equivalent to equipping each student with a tireless' personal coach ', providing precise one-on-one guidance that is impossible to achieve in traditional classrooms. This is the fundamental reason for the significant improvement in skill scores. Secondly, it promotes embodied cognitive learning. Sports learning is essentially a process of embodied cognition. This mode uses multimodal generation tools to transform abstract theoretical knowledge (such as mechanics principles) into visual information that students can intuitively see and closely associate with their own actions (such as action trajectory lines, force application diagrams), achieving the unity of physical experience and cognitive understanding, thus greatly promoting the learning effect in the cognitive field. Again, it nourishes the 'intrinsic motivation for self-determination'. According to Deci and Ryan's self-determination theory, this model satisfies students' "sense of ability" needs through immediate and transparent feedback; Satisfying the need for "autonomy" through personalized learning path selection; The gamified and contextualized learning design meets the need for "relatedness". The satisfaction of these three factors together gives rise to a strong intrinsic learning motivation, manifested in a significant increase in learning interest. Finally, it has initially cultivated 'advanced literacy for the future'. The improvement of deep learning ability is a pleasant surprise, but it is also reasonable. When students need to understand feedback reports from AI, engage in exploratory conversations with DeepSeek, and reflect on their learning process, they are actually engaged in critical thinking, knowledge integration, and metacognitive activities. This indicates that the value of the "DeepSeek+" model goes beyond the discipline of sports itself and has a positive impact on cultivating students' core competencies for the 21st century.

5.2 Practical insights and suggestions

Based on the above findings, the following suggestions are proposed: accelerate infrastructure and resource construction: Education administrative departments and schools should plan the construction of smart sports classrooms, integrate AI hardware and software platforms, and establish an AI resource library for sports teaching. This paper carry out teacher empowerment training: The core is to enhance teachers' "TPACK-AI" ability to proficiently use AI tools and redesign student-centered teaching activities. Clarify the ethical boundaries of human-machine collaboration: the principle of "teacher led" must be adhered to. Teachers are the designers of teaching, emotional caregivers, and ultimate reviewers of AI decisions. This study establishes a data privacy protection mechanism and educates students to critically view the content generated by artificial intelligence.

6. Conclusion and future work

This study draws the following conclusion through rigorous empirical testing: by systematically integrating the DeepSeek big language model and multivariate generative intelligence tools into physical education teaching, the "DeepSeek+" innovative teaching model constructed can produce statistically significant and practical educational improvements in teaching effectiveness. This model not only has excellent effects on students' mastery of sports skills and understanding of sports theory knowledge, but also effectively stimulates their interest in sports learning and preliminarily demonstrates the potential to promote the development of students' deep learning abilities.

Firstly, the sample comes from a school, and in the future, larger scale sampling can be conducted in different regions and student groups to test the universality of the model. Secondly, the experimental period is one semester, and its long-term effects (such as the impact on students' lifelong sports awareness) need to be tracked and studied. Again, this study mainly focuses on the effectiveness of the student end, and in the future, further research can be conducted on the impact of this model on teachers' professional development.

Looking ahead to the future, with the continuous evolution of multimodal modeling capabilities and their deep integration with the Internet of Things and virtual reality technology, the form of physical education classrooms will further evolve. We may usher in a "metaverse sports classroom" that combines reality and virtuality, where students can engage in unlimited tactical exercises with AI generated opponents in virtual space and gain in-depth biomechanical analysis based on full body motion capture. This study is only a small step towards future smart physical education, but it is a solid step with empirical support.

Acknowledgments

This work was supported by the Foshan Education Science "14th Five Year Plan" Special Project on Empowering Teacher Professional Development with Artificial Intelligence (Project No. 2025AIgh250).

ISSN 2618-1576 Vol. 7, Issue 5: 95-101, DOI: 10.25236/FSR.2025.070515

References

[1] Zhu Z T, Hu J. The Practical Logic and Promotion Strategy of Digital Transformation in Education [J]. China Electronic Education, 2022, (5): 1-10.
[2] Li Y. Research on Artificial Intelligence based Physical Education Action Recognition and Evaluation System [J]. Sports Science, 2021, 41 (8): 75-82.
[3] Wang Z J. How Generative Artificial Intelligence Reshaps Education: Current Status, Path, and Challenges [J]. China Electronic Education, 2023, (10): 1-10.
[4] Kirk D. Educational Value and Models-Based Practice in Physical Education[J]. Educational Philosophy and Theory, 2013,45(9), 973-986.
[5] Ryan,R.M.,&Deci, E.L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being[J]. American Psychologist,2000, 55(1),68-78.
[6] Nelson Laird, T. F., et al. Deeply Affecting First-Year Students' Thinking: Deep Approaches to Learning and Three Dimensions of Cognitive Development[J]. The Journal of Higher Education, 2014, 85(3), 402-432.
[7] Yin Z H. The dilemma and solution of sports learning evaluation under the background of core literacy [J]. Sports Journal, 2020, 27 (1): 121-127.