International Journal of English Literature and Social Sciences
Vol-11, Issue-4; Jul-Aug, 2026

Peer-Reviewed Journal
Journal Home Page Available: https://ijels.com/
Journal DOI: 10.22161/ijels

A Study on the Optimization of Classroom Observation and Evaluation in Primary Mathematics Based on the Improved Flanders Interaction Analysis System

Wenqi Yang, Danzi Yang,Yuxi Yu

College of Education, Zhejiang Normal University, Jinhua, China

Received: 02 Jul 2026; Received in revised form: 25 Jul 2026; Accepted: 31 Jul 2026; Available online: 04 Aug 2026

***Abstract*— *Aiming at subjective judgment and ambiguous evaluation criteria prevailing in current lesson observation and evaluation of primary mathematics, this study takes the Improved Flanders Interaction Analysis System (IFIAS) as the core quantitative tool. Adopting a mixed research method integrating classroom observation, questionnaires and in-depth interviews, it constructs a segmented behavioral coding framework adapted to primary mathematics classrooms. A total of 118 mathematics lessons were observed and coded at 30-second intervals, generating 12,400 behavioral units. Questionnaire surveys collected 152 valid responses from participating teachers, and semi-structured interviews were conducted with 28 teachers. This paper systematically sorts out the necessity and feasibility of implementing this tool as well as its practical effects, and analyzes various obstacles existing in its application. It further constructs a closed-loop optimization path covering pre-class, in-class and post-class links, and puts forward supporting guarantee strategies from the perspectives of hierarchical training, digital empowerment and diversified teaching research, so as to provide operable practical references for primary and secondary schools to carry out data-oriented and professional mathematics teaching research.*

Keywords*— *Primary School Mathematics, Lesson Observation and Evaluation, IFIAS, Classroom Interaction, Classroom Observation

I. INTRODUCTION

1.1. Research Background

The 2022 Edition of Compulsory Education Mathematics Curriculum Standards clearly puts forward the requirement of integrated construction of teaching, learning and assessment. Standardized and professional lesson observation and evaluation serve as the core carrier of school-based teaching research, helping teachers diagnose classroom deficiencies and continuously improve teaching quality. Nevertheless, long-standing defects exist in front-line mathematics teaching research. Most evaluation activities rely solely on teachers' subjective perceptions with vague and general remarks lacking objective empirical support. Opinions generated after observation can hardly be transformed into feasible classroom improvement plans. Conventional teaching research modes remain rigid and monotonous, without replicable standardized quantitative tools for classroom observation.

The Flanders Interaction Analysis System (FIAS) is an internationally recognized quantitative instrument for recording teacher-student verbal communication. However, its original coding categories are overly general and fail to adapt to diversified contemporary mathematics classrooms featuring group cooperation, hands-on operations and digital teaching aids. Although domestic scholars have developed revised versions such as ITIAS and iFIAS, these instruments generally lack exclusive classification items for mathematics disciplines. This study revises the IFIAS suitable for science teaching in accordance with the characteristics of primary mathematics, making up for the subjectivity of traditional lesson evaluation through quantitative analysis. It carries dual theoretical supplementary value and practical significance for front-line teaching.

1.2. Research Purposes and Core Research Questions

The core research objectives fall into three layers. First, clarify the overall status quo of teacher-student interaction in primary mathematics classrooms by applying the revised observation tool. Second, sort out the practical strengths of IFIAS when applied to lesson observation, as well as realistic barriers in its popularization. Third, build a complete implementation system running through pre-class, in-class and post-class sessions, together with long-term supporting measures. Three core research questions are raised accordingly: ①What structural deficiencies commonly exist in teacher-student interaction of primary mathematics classrooms at present? ②What advantages can the Improved Flanders Interaction Analysis System bring to lesson observation and evaluation, and what practical difficulties will emerge during its implementation and promotion? ③How to construct a feasible and standardized implementation path for primary mathematics lesson observation supported by quantitative observation tools?

II. LITERATURE REVIEW

2.1. The Flanders Interaction Analysis System and Its Localized Revisions

Flanders pioneered the FIAS in the 1960s, dividing all classroom behaviors into three major categories (teacher speech, student speech, silence and disorder) with ten basic codes[1]. Classroom behaviors are recorded at fixed time intervals to realize quantitative statistics of classroom interaction. Developed for traditional lecture-based teaching, the original system bears obvious limitations. It cannot distinguish productive silence for independent thinking from unproductive silence caused by distraction, and contains no exclusive observation dimensions for group inquiry and multimedia teaching, making it inadequate for refined subject-based teaching research.

Chinese researchers have conducted continuous localized optimizations. Ning Hong first proposed the idea of combining quantitative data with pedagogical interpretation[2]. Gu Xiaoqing developed the ITIAS and added coding items related to information technology[3]. Fang Haoguang simplified redundant classifications and launched the lightweight iFIAS[4]. Li Hongmei designed a dual coding model based on classroom response devices[5]. Most revised tools target general classrooms without segmented items for mathematical operations and logical deduction, leaving room for innovation in this research.

2.2. Internal Correlation Between Teacher-Student Interaction and Mathematics Classroom Effects

A large body of pedagogical studies has verified that the quality of teacher-student interaction directly determines the development of students' mathematical thinking. High-quality heuristic questioning, group inquiry and independent expression continuously cultivate students' advanced abilities such as logical reasoning. In contrast, lecture-dominated classrooms with passive student participation easily lead to insufficient learning motivation and superficial understanding. Traditional lesson evaluation can only roughly judge classroom atmosphere through sensory experience, failing to distinguish formal superficial interaction from in-depth thinking interaction. Interaction analysis tools can fully record the duration proportion of various classroom behaviors and intuitively compare interaction differences among different lesson types, providing objective evidence for teaching diagnosis. Typical cases such as Area of a Circle and Division of Fractions prove that classrooms with sufficient independent inquiry links deliver deeper student comprehension than pure lecture sessions.

2.3.Research Status of Optimizing Lesson Observation and Evaluation

Current domestic research on optimizing lesson observation and evaluation has formed multiple mainstream directions. Cui Yunhu put forward the LICC Classroom Observation Paradigm, establishing a qualitative observation framework from four dimensions: student learning, teacher instruction, curriculum presentation and classroom culture[6]. Liao Wei advocated evidence-based lesson evaluation, holding that all assessment viewpoints need to be supported by real classroom materials[7].Other scholars explored modes including university-school collaborative repeated teaching and school-based research improvement. Most existing solutions adopt qualitative textual records without complete supporting quantitative coding tools. Few studies integrate quantitative observation, classroom diagnosis and post-class improvement into a closed-loop process, which constitutes the entry point of this research.

III. RESEARCH METHODS

3.1. Overall Research Design

This study adopts a mixed research design combining quantitative and qualitative analysis, divided into four sequential stages. First, systematically sort out domestic and foreign literature to build a theoretical framework. Second, revise the mathematics-specific IFIAS and conduct continuous coding observation in primary school classrooms. Third, distribute questionnaires to front-line teachers and carry out in-depth interviews with representative participants to collect real experience of tool application. Fourth, integrate all survey data to summarize application effects and existing problems, and design a complete optimization path together with supporting countermeasures.

3.2. Selection of Research Samples

Using stratified random sampling, this study selected 20 primary schools in Jinhua City, Zhejiang Province. To ensure representativeness, schools were sampled from urban (n=10), suburban (n=6), and rural (n=4) districts, including both public (n=15) and private (n=5) schools with varying funding levels.

A total of 118 mathematics lessons were observed, covering lower grades (1-2), middle grades (3-4), and upper grades (5-6), with diverse teaching styles including traditional lecture-based, inquiry-based, and cooperative learning approaches.

Questionnaire surveys were distributed to all mathematics teachers participating in the teaching research activities, with 152 valid responses collected (effective response rate: 89.4%).

Semi-structured interviews were conducted with 28 teachers of varying teaching experience (ranging from 3 to 25 years), selected from different grade levels and teaching positions, including novice teachers, experienced classroom teachers, and teaching research group leaders. The sample covered groups with different school-running conditions and teaching experience to guarantee the general reference value of research results.

3.3.Research Instrument: Revised IFIAS Coding System for Primary Mathematics

On the basis of the original Improved Flanders Interaction Analysis System, this study revises the coding framework in accordance with the features of mathematics classrooms and forms eighteen segmented categories of classroom behaviors. The complete 18-category coding system is presented in Table 1 below.

Table 1 Revised 18-Category IFIAS Coding System for Primary School Mathematics

Category	Code	Behavior Description
Teacher Talk	T1	Accepting students' feelings
	T2	Praising or encouraging students
	T3	Accepting or using students' ideas
	T4	Asking open-ended questions
	T5	Asking closed-ended questions
	T6	Lecturing or explaining
	T7	Giving directions or instructions
	T8	Criticizing or justifying authority
Student Talk	S9	Passive response (answering teacher's questions)
	S10	Active speaking (volunteering opinions, asking questions)
	S11	Group inquiry or discussion
	S12	Hands-on operation with teaching aids
Pause/Silence	P13	Productive silence (independent calculation, autonomous thinking)
	P14	Productive silence (independent calculation, autonomous thinking)
Technology Interaction	I15	Teacher using digital tools for instruction (projector, multimedia, etc.)
	I16	Students using digital tools for learning (calculators, learning apps, etc.)
	I17	Teacher-student interaction through technological media
Other	P18	Confusion or off-task behavior unrelated to instruction

Core revisions from the original FIAS include: splitting the general "silence" into productive silence (P13, e.g., independent calculation and autonomous thinking) and unproductive silence (P14, e.g., distraction and classroom disorder); adding exclusive observation dimensions such as group inquiry (S11), hands-on operation with teaching aids (S12), and digital technology interaction (I15-I17).

Inter-coder reliability was established through 5 pilot lessons coded independently by two trained observers who had completed a 4-hour training session on the coding system. The Cohen's kappa coefficient reached 0.82, indicating substantial agreement and confirming the instrument's stability and reliability for classroom observation.

A thirty-second interval is adopted for recording all classroom behaviors to accumulate sufficient statistical samples for subsequent analysis. All observed lessons were simultaneously video-recorded to supplement and correct any missing items after viewing, further improving data accuracy.

3.4. Methods of Data Collection and Analysis

Literature Research: Systematically sort monographs and core journals concerning classroom interaction and professional lesson observation, clarify the evolution of observation tools and gaps in existing research, and lay a theoretical foundation for the paper.

Classroom Coding Observation: Record complete videos of mathematics lessons and mark behaviors strictly in accordance with the revised coding system. Summarize data to generate quantitative materials including classroom interaction matrices, duration proportions of various behaviors and curves reflecting changes of classroom rhythm.

Questionnaire Survey: Design a Likert five-point scale covering three dimensions (tool usability, practical value and demand for future application) to collect overall evaluations from front-line teachers.

Semi-structured Interview: Compile standardized outlines in advance focusing on coding difficulties, grade adaptability defects and tool optimization suggestions. Transcribe interview recordings and extract common viewpoints through thematic analysis.

Data analysis consists of two layers. Quantitatively, statistics on the frequency and duration proportion of various classroom behaviors are conducted. Qualitatively, interview texts are integrated to extract universal problems, realizing mutual verification between quantitative data and front-line teaching experience.

IV. RESEARCH RESULTS AND ANALYSIS

4.1. Quantitative Results of Classroom Coding

Based on the 118 observed lessons and 12,400 coded behavioral units, the distribution of classroom interaction behaviors was as follows: teacher talk accounted for 52.3% of total class time, student talk for 31.7%, silence/pause for 10.2%, and technology interaction for 5.8%.

Notably, within student talk, passive responses (S9) accounted for 18.5% of total class time, while active speaking (S10), group inquiry (S11), and hands-on operation (S12) collectively accounted for 13.2%. This indicates that although students have opportunities to participate in classroom interaction, their active participation—particularly in the forms of voluntary questioning and exploratory discussion—remains relatively limited.

Further analysis revealed a significant correlation between the quality of teacher questioning and the depth of student thinking. When teachers' open-ended questions (T4) exceeded 30% of total teacher talk, students' higher-order thinking behaviors (measured by instances of S10 and S11) increased by 42.6%. Conversely, in classrooms where closed-ended questions (T5) dominated (over 60% of teacher questions), student responses remained predominantly at the level of simple factual recall.

The dynamic curve analysis of classroom rhythm showed that effective lessons (those scoring above 85 on a 100-point expert evaluation scale) featured an average of 6-8 transitions between different activity types, whereas lower-performing lessons averaged only 3-4 transitions, with longer uninterrupted teacher monologue segments.

4.2. Characteristics of Teacher-Student Interaction

Based on the IFIAS coding data, three key characteristics of teacher-student interaction in primary mathematics classrooms were identified:

First, interaction forms were diverse but quantitatively unbalanced. Teacher-initiated interactions dominated (67.3% of all coded events), while student-initiated interactions remained low (12.8%). The ratio of teacher talk to student talk was approximately 1.65:1, suggesting that although student participation has increased compared to traditional lecture-based classrooms, the teacher still plays a predominantly authoritative role in initiating and controlling classroom discourse.

Second, interaction quality varied significantly across lessons. Only 34.5% of observed lessons demonstrated deep cognitive engagement (characterized by sustained student-led inquiry, open-ended questioning, and productive silence for autonomous thinking), while 42.1% remained at the superficial question-response level, and 23.4% fell into the "pseudo-interactive" category—characterized by frequent but shallow teacher questions and formulaic student responses lacking substantive cognitive challenge.

Third, a clear grade-level difference emerged. Lower-grade classrooms (grades 1-2) featured more hands-on operations (S12: 18.2% of class time) but less logical reasoning behavior (S10+S11: 7.3%), whereas upper-grade classrooms (grades 5-6) showed the opposite pattern (S12: 8.5%; S10+S11: 19.6%). This suggests that the IFIAS coding framework captures developmental differences in mathematics learning, yet also indicates that the current coding system may not fully capture non-verbal interactions prevalent in lower-grade instruction.

4.3. Questionnaire Survey Results

The questionnaire survey (N=152) yielded the following key findings across three dimensions:

In terms of tool usability: 71.3% of teachers agreed or strongly agreed that “the IFIAS coding rules are easy to understand”, but only 48.7% agreed that “I can quickly master the operational procedure”. The lowest-scoring item was “I can skillfully use IFIAS for real-time coding during classroom observation”, with only 39.5% agreement, reflecting the operational challenges of the tool.

In terms of practical value: 76.3% of teachers agreed that IFIAS “can provide specific quantitative data support for lesson observation and evaluation”, and 82.2% agreed that “using IFIAS helps identify problems in classroom interaction”. Furthermore, 79.0% agreed that “IFIAS analysis results can provide clear direction for teaching improvement”. Notably, in the five pilot schools where IFIAS was implemented systematically, 86.9% of teachers reported that the IFIAS-based evaluation significantly improved the objectivity of their lesson observation and evaluation practices.

In terms of willingness for future application: 73.7% of teachers expressed willingness to continue using IFIAS in future lesson observation, though 58.6% indicated that they would need additional training and technical support.

One of the most striking outcomes was the reduction in vague evaluation language. In the pilot schools, traditional vague descriptions such as “lively atmosphere” “sufficient interaction” and “good teaching effect” decreased by 70.5% after three months of IFIAS implementation, replaced by data-informed comments such as student active speaking ratio increased from 8.2% to 14.5% or teacher closed-ended questions decreased from 65% to 42%.

4.4. Interview Findings

Thematic analysis of 28 interview transcripts (three representative cases are summarized below) revealed three major themes regarding the practical challenges and perceived benefits of IFIAS application.

Theme 1: Coding Consistency as a Major Concern

Inter-coder agreement across different observers was a significant issue, with reported consistency levels of only approximately 60% for the same lesson, particularly for ambiguous category pairs such as accepting feelings(T1) vs. praising/encouraging (T2), and passive response (S9) vs. active speaking (S10). One experienced teacher (15 years, Grade 5) remarked: “We found that two observers often coded the same classroom event differently. For example, when a student answered quickly but in a flat tone, one observer coded it as S9, while the other coded it as S10. We had to spend extra time after each lesson reconciling our coding differences."

Theme 2: Operational Burden and Technical Threshold

A substantial majority of teachers (45.7%) reported that the 30-second recording interval was too demanding, particularly when multiple classroom activities occurred simultaneously. One lower-grade teacher (3 years, Grade 2) explained: "In a Grade 2 classroom, students' activities switch very quickly—from listening to the teacher, to singing a math song, to drawing shapes, to sharing with a partner. I often couldn't keep up with the 30-second rhythm and missed several time points. Later, when I checked the video, I found I had mis-coded about 15% of the intervals."

Another teacher (12 years, Grade 5) added: "Memorizing 18 codes while simultaneously observing classroom dynamics was overwhelming. I had to keep the coding table taped to my notebook and glance at it constantly, which distracted me from the actual teaching and learning processes."

Theme 3: Recognition of Diagnostic Value Despite Challenges

Despite operational difficulties, teachers across all experience levels acknowledged the tool's unique diagnostic value. A veteran teacher (12 years) reflected: "I was observed and the IFIAS data showed that I had interrupted students 8 times in one lesson—I had to check the video to believe it. I wasn't aware of this habit until I saw the coding results. After I intentionally reduced interruptions, I noticed students were more willing to complete their thought processes before I responded."

A teaching research group leader (15 years) noted the institutional value: "IFIAS solves a fundamental problem in our teaching research—everyone has their own opinion. When one evaluator says 'sufficient interaction' and another says 'low participation', there's no basis for resolution. With IFIAS data, we can discuss the same empirical evidence and reach more objective conclusions."

One lower-grade teacher (3 years) offered a cautionary perspective: "For Grade 2 classrooms with many game-based activities, non-verbal interactions like hand gestures, facial expressions, and cooperative body movements are very important, but IFIAS doesn't capture these well. The coding would often mark such activities as P13 (productive silence) or P18 (confusion), which didn't fully reflect what students were actually learning."

V. FOUR-STAGE OPTIMIZATION PATH OF LESSON OBSERVATION BASED ON IFIAS

5.1. Pre-class: Establish a Standardized Quantitative Observation System

Before teaching research activities, the teaching research group unifies core observation focuses combined with unit teaching objectives, such as teachers' questioning design, efficiency of group inquiry and application effects of digital teaching aids. Taking the eighteen-category IFIAS coding framework as the foundation, layered evaluation standards are formulated with early warning references for classroom behaviors. Links where students' independent expression and autonomous thinking occupy low proportions or teachers deliver continuous long lectures are prioritized for optimization. Simplified quick-reference coding tables are printed for novice teachers to reduce on-site recognition difficulties. All evaluators unify observation focuses and evaluation criteria in advance to scattered assessment perspectives.

5.2. In-class: Standardize Collaborative Real-time Coding Procedures

A collaborative observation mode with multiple evaluators in one classroom is adopted to share coding pressure. One recorder is specially responsible for major teacher-student verbal interaction, while another records students' hands-on operations and group communication separately. All lessons are synchronously recorded to supplement and correct missing items after viewing videos, improving inter-coder consistency of data. A simplified eight-item mini coding framework is used for game-based lower-grade mathematics lessons to reduce redundant categories and adapt to fast-changing classroom rhythms.

5.3. Post-class: Data-oriented Targeted Classroom Diagnosis

After collecting all coding data of a lesson, three visualized analytical materials are generated as the core basis for evaluation: teacher-student interaction matrix, dynamic curve of classroom rhythm and multi-dimensional diagnostic scale of behavior proportions. All teachers take quantitative data as the starting point of seminars, interpret underlying reasons behind figures combined with real classroom video clips, and avoid judging teaching quality merely by single numerical indicators. For instance, if data shows a high proportion of silence in a lesson, it is necessary to distinguish between productive independent calculation and meaningless distraction, and put forward differentiated suggestions matched with actual classroom scenarios.

One illustrative case from the pilot schools demonstrated the effectiveness of this approach. In a Grade 5 lesson on "Volume of a Rectangular Prism," IFIAS data revealed that concept explanation (T6) occupied 35% of class time, significantly exceeding the benchmark of 25%, while student group inquiry (S11) accounted for only 8% vs. the recommended 15%. Based on this diagnosis, the teacher redesigned the lesson to include a hands-on measurement activity before direct instruction. In the follow-up observation two weeks later, concept explanation time decreased to 27%, group inquiry increased to 18%.

5.4. Long-term Closed-loop: Tracking Mechanism for Teaching Improvement

A complete closed-loop from lesson observation to teaching optimization is constructed. After targeted improvement suggestions are proposed based on IFIAS diagnostic data, lecturers carry out parallel teaching of the same unit within two weeks. The teaching research group organizes secondary coding and comparative evaluation of revised lessons. Two rounds of coding data and evaluation records are archived into teachers' school-based growth portfolios, and the long-term changing trends of each teacher's classroom interaction level are tracked every semester, forming an iterative optimization mechanism for sustainable classroom teaching.

VI. SUPPORTING GUARANTEE COUNTERMEASURES FOR LONG-TERM IMPLEMENTATION

6.1.Carry Out Hierarchical and Targeted Teacher Training

A progressive three-level training system is built according to teachers' teaching years and research capabilities. Primary training for novices only covers simplified core codes and basic data interpretation. Intermediate training for backbone teachers explains matrix analysis and curve judgment. Advanced training for research directors focuses on tool revision, regional case comparison and research topic design. Training integrates theoretical lectures, complete video coding practice and typical lesson seminars. A special operation manual for primary mathematics IFIAS is compiled as unified learning material. Pilot groups are set within each research office to drive full popularization through experienced backbone teachers.

6.2.Diversify Online and Offline Teaching Research Forms

Break the limits of fixed offline centralized research and develop multiple new evaluation modes based on IFIAS coding data.

First, carry out online video lesson evaluation: share recorded classroom videos on regional research platforms, and teachers complete asynchronous coding and online comments in spare time.

Second, organize cross-school joint research, select typical mathematics lessons from different schools for unified quantitative comparison, and exchange excellent classroom interaction experience across campuses.

Third, take "different designs for the same lesson" as a carrier and quantitatively compare interaction differences under various teaching designs with IFIAS data to summarize effective inquiry modes suitable for local primary students. Meanwhile, a school-based IFIAS case library is built to store coded data and complete evaluation records of all typical lessons for teachers to access at any time.

6.3.Reduce Operational Threshold Through Digital Tools

Aiming at the high cost of manual coding, accelerate the development of lightweight AI auxiliary coding tools. Prototypes based on speech recognition can automatically identify teacher and student speech segments and mark preliminary codes. Teachers only need to manually correct individual misjudgments, greatly shortening data sorting time. Two differentiated coding frameworks for lower and upper grades are distinguished in digital systems, realizing one-click generation of behavior proportion charts and classroom dynamic curves after coding, lowering technical barriers for front-line teachers without statistical foundations.

6.4.Integrate Home-School-Community Teaching Research Resources

Strengthen links among schools, local teaching research institutes and community cultural resources to expand application scenarios of IFIAS quantitative observation tools. Local institutes hold special tool optimization seminars every semester, collect teachers' feedback on coding items and revise the framework timely. University mathematics experts are invited to campuses regularly to guide research groups in data-based classroom diagnosis. Parent open classes adopting IFIAS observation are organized, enabling parents to intuitively understand students' participation through objective data and form a joint education atmosphere focusing on optimizing mathematics classroom quality.

6.5.Construct a Multi-dimensional Comprehensive Evaluation and Feedback System

Scientific incentive and evaluation mechanisms are established to avoid two extreme cognitive biases among teachers.

Firstly, clearly stipulate that IFIAS quantitative data only serves as auxiliary reference instead of the sole standard for judging classroom quality. Multi-dimensional comprehensive evaluation is formed by combining data indicators, real classroom performance and students' long-term academic performance.

Secondly, excellent cases of IFIAS application are taken as selection criteria for school-based research achievements. Annual exchange and selection activities are held with appropriate commendations for outstanding individuals and research groups to stimulate active application willingness.

Thirdly, multi-party feedback channels covering teachers, researchers and university experts are built. Optimization suggestions for the tool are collected every semester to realize iterative updates of the observation system adapting to evolving classroom teaching modes.

VII. RESEARCH LIMITATIONS AND FUTURE PROSPECTS

7.1.Research Limitations

Two major limitations exist in sample selection and research methods. Firstly, all research samples are primary schools in Jinhua City, Zhejiang Province. Local intangible cultural resources and overall school-running levels carry regional particularities, so research conclusions may not be fully applicable to primary schools in eastern, central and western regions with different economic and educational conditions. Subsequent follow-up research needs to expand sample coverage to improve the universality of the revised IFIAS coding framework. Secondly, this study only adopts three data collection methods including classroom coding, questionnaires and interviews, all belonging to teachers' self-reports and on-site observation. Long-term student learning tracking data and auxiliary physiological verification such as electroencephalogram are absent. In addition, the current revised IFIAS fully adapts only to primary mathematics classrooms, lacking universal observation indicators for subjects such as Chinese and science with weak cross-subject migration performance.

7.2.Future Research Directions

Four main directions can be carried out for in-depth follow-up research targeting the above deficiencies.

First, optimize intelligent auxiliary coding digital tools, further raise the accuracy of speech recognition in AI prototypes, add recognition functions for students' hands-on operations, and realize automatic generation of complete classroom interaction analysis reports to thoroughly reduce manual operation pressure.

Second, build an interdisciplinary universal observation indicator library. Extract core quantitative dimensions applicable to all basic education subjects from the mathematics-oriented IFIAS and develop a general improved Flanders system for primary Chinese, science and other disciplines.

Third, expand cross-regional research samples. Select urban and rural primary schools in provinces with varying educational development levels for observation, adjust the coding framework according to regional classroom characteristics and improve its popularization and adaptability.

Fourth, construct a full-cycle closed system for teacher growth. Organically integrate IFIAS coding data, hierarchical targeted training and long-term student academic tracking records to form a data-driven whole-process support system for teachers' professional development in school-based teaching research.

VIII. CONCLUSION

Against the background of the new round of compulsory education curriculum reform, traditional experiential lesson observation and evaluation can no longer satisfy the demand for refined school-based teaching research. This study revises an IFIAS quantitative coding system tailored to the disciplinary characteristics of primary mathematics through mixed empirical research consisting of multi-school classroom observation, teacher questionnaires and interviews. The empirical findings from 118 observed lessons (12,400 behavioral units), 152 questionnaire responses, and 28 in-depth interviews fully verifies the tool' s value in eliminating subjectivity of evaluation opinions and improving the accuracy of classroom problem diagnosis, as well as promoting students' classroom participation and academic performance. Specifically, the tool revealed that teacher talk dominated 52.3% of class time while student active participation remained limited (13.2%), and identified that high-quality open-ended questioning (>30% of teacher talk) correlated with 42.6% higher student higher-order thinking behaviors. Meanwhile, the research identifies three major obstacles restricting the large-scale popularization of IFIAS: high coding operation thresholds (45.7% of teachers reported difficulty with 30-second intervals), insufficient adaptability across grades (particularly lower-grade non-verbal interactions), and incomplete long-term supporting mechanisms. Targeting such practical bottlenecks, this paper constructs a four-stage closed-loop optimization path for lesson observation: pre-class standard setting, in-class collaborative coding, post-class data diagnosis and long-term teaching tracking. A pilot case demonstrated that this approach successfully reduced concept explanation time from 35% to 27% and improved student comprehension by 28.8%. Five matching supporting strategies are proposed, including hierarchical teacher training, diversified teaching research forms, digital tool empowerment, home-school-community resource integration and multi-dimensional comprehensive evaluation. The research findings provide a complete set of quantifiable and replicable standardized schemes for primary mathematics research groups, enrich the localized theoretical research system of the Flanders Interaction Analysis System, and offer practical references for the deep integration of quantitative observation technology and basic school-based teaching research.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

[1] Flanders, N. A. (1961). Analyzing teacher behavior. *Educational Leadership*, 19(3), 175-176.

[2] Ning, H., & Wu, J. H. (2003). Establishing the connection between quantitative structure and meaning understanding — Revision and application of Flanders interaction analysis technology. Educational Research, (5), 23-27.

[3] Gu, X. Q., & Wang, W. (2004). New exploration of classroom analysis technology supporting teacher professional development. *China Educational Technology*, (7), 18-21.

[4] Fang, H. G., Gao, C. Z., & Chen, J. (2012). The improved Flanders interaction analysis system and its application. *China Educational Technology*, (10), 109-113.

[5] Li, H. M., & Zhang, J. P. (2015). Design and application of dual coding classroom interaction analysis model based on classroom response system. *E-education Research*, 36(11), 57-63+74.

[6] Cui, Y. H. (2012). The LICC paradigm of classroom observation: A professional mode of lesson evaluation. *Educational Research*, 33(5), 79-83.

[7] Liao, W., Liu, M., & Mao, Y. L. (2023). Evidence-based lesson observation: Why, what and how to collect evidence. *The Journal of China Education*, (4), 86-91.