A Study on Diagnosis and Improvement Path of Outdoor Play Guidance Skills for Early Childhood Teacher Trainees Based on Multi-Source Data Fusion

Laling Dong
Academic Affairs Office, Jinan Preschool Education College, Jinan 250000, Shandong, China
Corresponding author: Laling Dong (e-mail: LalingDong_1989@163.com)

ABSTRACT Outdoor play guidance in early childhood education involves rapid perception of children's behavior, continuous visual attention allocation, and stable decision-making under uncertain outdoor conditions. To improve the objectivity and real-time capability of skill diagnosis for early childhood teacher trainees, this study proposes a multi-source data fusion framework integrating video-based behavior analysis, eye-tracking, physiological signal monitoring, and questionnaire data. A multi-angle camera system is used to capture guidance behaviors and movement trajectories, Tobii Pro eyetracking data characterize visual scanning and hazard attention, and wearable devices record heart-rate-variability indicators reflecting stress regulation. After temporal synchronization, feature normalization, and adaptive weighting, K -means clustering and an SVM classifier are employed to identify guidance-skill levels. Experimental analysis of 180 teacher trainees shows that the high-level group exceeds the low-level group by 27.4 percentage points in effective observation rate, and the SVM model reaches 87.6% classification accuracy with strong agreement with expert ratings. The proposed framework provides an engineering-oriented approach for multimodal sensing, physiological signal processing, and intelligent skill diagnosis in complex human–environment interaction scenarios.

INDEX TERMS multi-source data fusion, eye tracking, physiological signal processing, visual sensing, skill diagnosis

I. INTRODUCTION

KINDERGARTEN outdoor play offers children a valuable environment where they can enjoy nature as well as the motor and social skills and be able to form social relationships. The achievement of the educational value of play directly depends on the quality of guidance provided by teachers. The outdoor play guidance ability training of the preschool teacher in the present training system is not coordinated, because there are no systematic assessment tools to evaluate the sensitivity of the observation, the time of intervention and the ability to react to an emergency. Teacher trainees frequently exhibited a dichotomy in their approach, oscillating between over-intervention and complete neglect by some teacher trainees during their internships in which they manifested conspicuous stress responses and indecision in decision making in case of emergency situations like children climbing and falling or clash of materials. Conventional methods of evaluation are based on subjective performance of the supervising teachers which is constrained by the subjective views on observation and time and energy, and as such, it may not be able to fully describe the dynamic nature of guidance behavior [1,2]. The absence of objective and quantitative diagnostic methods results in poor targeting of training programs, which makes it hard to supply teacher trainees with the relevant ability feedback and paths of improvement, thereby limiting the quality of initial professional development of preschool teachers. As a new level of reference, multi-source data fusion technology perspective to evaluate teacher competence assessment. Video analysis method realizes behavioral trajectory and interaction frequency quantification, eye tracking method reflects the cognitive process of the interaction behavior allocation, physiological signal monitoring method realizes objective change of emotional arousal and stress level, questionnaire and interview method realizes subject experience and in-depth information of reflection. Machine learning method realizes the extraction of discriminative feature from high dimensional heterogeneous data, builds automatic skill level identification model, and breaks the subjectivity and low efficiency of manual assessment. This paper takes four kinds of data source into account, builds a diagnostic based outdoor play guidance skills of preschool teacher

trainings, uses K-means clustering and SVM classification algorithm accurately identification skill level, designs a virtual simulation and real practice training program based on the diagnostic results, verifies the effect of data-driven personalized intervention, and provides technical support and empirical basis to improve the quality of preschool teacher education [3,4].

II. RELATED WORK

The study of the capability of early childhood teachers to facilitate outdoor play has been developed not only in theory but also quantitatively. Initial works concerned the explanation of the role of teachers and intervention principles in play with the emphasis on the need to observe children interests, respond to play needs, and provide a positive environment but did not micro-analyze and quantitatively assess guidance behaviors. Over the last few years, researchers started to address the observation and interpretation skills, the choice of support strategies, and the level of reflective practice of teachers and create assessment instruments which involve the aspects of records of observations, quality of interactions and professional judgments. Existing scales, however, are largely based on self-report or peer-report measures, thus challenging to measure the immediate decision-making process and implicit cognitive properties in real world scenarios [5,6]. The multi-source data fusion technology has proven to be beneficial in incorporating heterogeneous information to enhance the accuracy of decision making in medical diagnoses, testing in industries and financial risk management. Its use in education is slowly coming out. Scholars have tried to merge video recordings in the classroom, learning management system records and physiological sensor data to assess the effectiveness of teaching, where multimodal data can provide more intricate patterns that cannot be indicated by one source of data. Attention allocation by the teacher towards the behavior of the students has been studied using eye-tracking technology and results reveal that the expert teachers possessed a larger range of visual scanning and faster rate of recognizing an abnormal behavior. The indicators of heart rate variability measured by wearable gadgets have been demonstrated to be useful in assessing the level of stress and the emotional regulation capacities of teachers in managing classrooms. Machine learning algorithms used in evaluating the professional competence of teachers involve cluster analysis to determine the types of teaching style, classification models to determine the level of teaching effectiveness, and sequence mining to determine the patterns of teaching behavior that are effective. Nevertheless, the available studies concentrate on classroom teaching scenarios and there is still no use of technology to instruct children in outdoor activities. In addition, there is no theoretical direction in the approach of data fusion, which makes the weight of features arbitrarily set [7,8].

III. METHODS

A. RESEARCH DESIGN AND SUBJECT SELECTION

This research used a mixed-method research design, where 180 third-year teacher-training students in five universities in East China were used as the research subjects, majoring in preschool education. Participants were recruited through stratified random sampling from the colleges' internship databases to ensure a representative cross-section of the regional trainee population. There were 168 female and 12 male among them, the mean age of which was 20.8 years. While the gender imbalance (93.3% female) reflects the current demographic reality of the preschool education field in China, preliminary T-test analysis showed no statistically significant differences in baseline skill scores between male and female participants ($p > 0.05$), suggesting that the diagnostic model's findings are broadly applicable across genders within this professional context. They were all presently enrolled in core courses like psychology of preschool child development and kindergarten curriculum and teaching, but had not yet been systematically trained in the guidance of outdoor play. The experimental locations were outdoor activity areas in three kindergartens, with four common play areas, such as sand and water areas, climbing areas, role-playing areas and nature exploration areas. All teacher-training students had to instruct 4-6 year old children to play independently in a natural environment in 30 minutes and at the same time, gathering multimodal data. The trial period was predetermined to be 9:00 AM to 11:00 AM time so that the children would be in the same state of activities. To manage confounding factors, all the participants became acquainted with the site environment a week ahead of time, however, not notified about particular indicators of assessment [9,10]. This study was approved by the Ethics Committee of Jinan Preschool Education College, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form. The principles of child protection were adhered to in the data collection process, and the faces of children were distorted in all video materials. Table 1 presents the detailed distribution characteristics of the research subjects and the configuration information of the experimental site.

TABLE 1. Distribution of research subjects and configuration of experimental sites

University Number	Number of participants	Gender ratio (female/male)	Experimental Kindergarten	Game Region Type	Duration of a single guidance session
University A	42	39/3	Municipal Demonstration Park	Sand and water area + climbing area	30 minutes
B Normal College	38	36/2	Provincial Demonstration Park	Role-playing area + Nature exploration area	30 minutes
C Higher Vocational College	35	32/3	Ordinary public kindergartens	Sand and Water Area + Nature Exploration Area	30 minutes
D Vocational College	33	30/3	Inclusive private kindergartens	Climbing area + role-playing area	30 minutes
E Normal University	32	31/1	Municipal Demonstration Park	Full coverage of four types of areas	30 minutes
Total	180	168/12	3 kindergartens	Four typical spaces	5400 minutes

B. CONSTRUCTION OF MULTI-SOURCE DATA ACQUISITION SYSTEM

Eight Sony FDR-AX700 cameras were arranged for video behavioral data collection around the play area. The cameras were fixed at 45 degrees to the pillars around the play area to construct a 360 degree view around the play area without blind spots. The cameras resolution was recorded

at 1920 × 1080 and the frame rate was 30fps. The cameras were connected to a Dell Precision workstation via a local area network and the OpenPose posture recognition algorithm was implemented to annotate the body orientation, movement trajectory and the movement of the limbs of the student teachers in real time. Eye-Tracking data were collected via a portable eye tracker, Tobii Pro Glasses 3, with a sampling frequency of 100Hz and pupil tracking accuracy of 0.5 degrees. The eye tracker was calibrated with a five-point calibration method and recorded the time length and switching frequency of student teachers gaze on three types of targets, namely single children, play materials and safety hazards. Physiological signals were collected with an Empatica E4 wristband sensor and sampled heart rate, skin conductance response and body surface temperature at a sampling frequency of 64Hz. The temporal index SDNN and frequency index LF/HF ratio of heart rate variability were calculated to reflect the emotional arousal level and stress response intensity of student teachers. Questionnaire interviews were implemented with a semi-structured game observation strategy scale and a self-assessment form for guided behavior which were filled in immediately after the guided game session with iPad. Four types of data sources were synchronized with timestamps to construct an integrated data warehouse.

C. DATA FUSION AND SKILLS DIAGNOSIS MODEL

After running OpenPose, video were further used to extract three behavioral features of teacher trainees, i.e., movement distance, dwell time and interaction frequency. The dimension of 25 skeletal keypoints was reduced into 8-dimensional motion pattern vector by autoencoder. Eye-tracking data were used to calculate the heatmap distribution entropy value of gaze point distribution to characterize the systematicness of gaze. Region of interest analysis was used to explore teacher trainees' sensitivity to the warning of dangerous behavior. Wavelet transform was used to filter motion artifacts from physiological signals. To further minimize the confounding effects of physical exertion on HRV and SCR, 3-axis accelerometer data from the E4 wristband were used as a reference to regress out motion-induced noise, ensuring the signals primarily reflected psychological stress and emotional arousal. RMSSD index, which represented the activity of parasympathetic nerve, was extracted as the feature of RR interval sequence. The feature of skin conductance response peak count represented the number of stress events by counting the number of peaks. Multi-source feature fusion uses an adaptive weighted strategy. The initial weights (video: 0.35, eye-tracking:

0.30, physiological: 0.20, questionnaire: 0.15) were determined through a combination of expert consultation (Delphi method) and a preliminary sensitivity analysis. Specifically, video and eye-tracking data were assigned higher weights as they directly capture observable guidance behaviors and cognitive focus, which expert practitioners identify as the primary indicators of instruction quality. To ensure the robustness of the fusion model, these weights were further optimized via 10-fold cross-validation on the training set to minimize classification error, ensuring that the 64-dimensional fused feature vector represents the most discriminative information across modalities. To establish a reliable ground truth for the skill levels, three senior kindergarten teaching experts (each with >15 years of experience) independently rated a subset of 50 participants' performance using a standardized rubric. The K-means clustering results (k=3) were then validated against these expert ratings, achieving a high degree of concordance (89.2% agreement). The SVM classifier was subsequently trained using these validated labelst. The parameter of penalty parameter C is tuned by grid search to be 100. The parameter of kernel parameter gamma is set to be 0.01.

D. DESIGN OF A THREE-STAGE PROGRESSIVE CULTURE PROGRAM

The scenario simulation training adopted the HTC Vive Pro virtual reality device to build four kinds of outdoor game scenarios. Teacher trainees wore the head-mounted display to enter the scenario and responded in real time to 12 typical events happening to virtual children, including climbing up and falling down, peer conflicts, and disputes over materials. The time of intervention and type of support strategies were recorded by controlling the operation of hands. Each training session lasted 45 minutes and could finish 8 scenario cycles.

The case study session played video recordings of expert guidance given by instructors. We annotated observation points, verbal content, and body language of student teachers and experts frame by frame in 5- second time windows with the video analysis software ELAN. We put student teachers in groups to discuss the reasoning of experts in making decisions and compared the differences in handling similar situations with students. Mentors facilitated in-depth reflection of student teachers using the Socratic questioning method. Each session lasted two hours and covered 6 case segments.

The arrangement of reflective practice asks teacher trainees to come back to the kindergarten for the second guidance session, wearing cameras portable for them to enter the scene from the first-person perspective, and watching the videos recorded by themselves after the guidance session, and then filling out a structured reflection log, which includes three modules: identification of critical incident, reconstruction of the decision process, and development of improvement strategy. One reflection report is required per week for 8 weeks.

IV. RESULTS AND DISCUSSION

A. RESULTS OF MULTI-SOURCE DATA FEATURE ANALYSIS

Descriptive statistics and ANOVA were executed on the multi-source data from 180 teacher trainees. Video data were used for the cumulative observation time of each participant in the guidance process of 30-minutes and calcu-

lated the percentage of effective observations. Eye-Tracking data were used for the total time of individual gaze at the children and the number of switching of gaze. Physiological data were used for the SDNN mean and LF / HF ratio to reflect the degree of stress. Questionnaire data were used for the summary of scores in the three dimensions of observation strategy, interaction, and reflection ability. One-way ANOVA was used to test the significance of the differences between the three skill level groups. The effect size was calculated by eta squared. Table 2. Comparison of characteristics of multi-source data from different skill level groups.

TABLE 2. Comparison of features of multi-source data for different skill level groups

Skill Level Group	Sample size	Percentage of valid observations (%)	Gaze switching frequency (times/minute)	SDNN mean (ms)	LF/HF ratio	Observation strategy score	Interaction quality score	Reflection ability score	Overall Skills Score
High-level group	52	68.7±6.3	8.4±1.2	54.8±7.6	1.8±0.4	4.3±0.5	4.5±0.4	4.2±0.6	83.6±5.8
Intermediate level group	76	54.2±8.1	5.7±1.5	42.3±9.2	2.6±0.6	3.5±0.6	3.6±0.5	3.4±0.7	68.4±7.2
Low level group	52	41.3±9.5	3.2±1.8	28.1±8.4	4.1±0.9	2.7±0.7	2.8±0.6	2.5±0.8	51.2±8.6
F value	-	186.34***	152.78***	141.62***	198.45***	124.56***	168.92***	132.73***	312.87***
p-value	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
eta^2	-	0.678	0.633	0.615	0.691	0.584	0.656	0.600	0.779

The high-level group differed from the low-level group by 27.4 percentage points in the percentage of effective observations, indicating that outstanding teacher trainees were able to maintain sustained attention to the children's play rather than being distracted. Their fixation switching frequency was 5.2 times per minute higher, reflecting stronger initiative and systematicity in visual scanning. Their mean SDNN time was 26.7 milliseconds higher, and their LF/HF ratio was only 43.9% of that of the low-level group, indicating that the high-level group maintained a lower level of sympathetic nerve activation and better emotional regulation

during guidance. The questionnaire scores in all three dimensions showed significant gradient differences. The variance analysis of the comprehensive skill score had an F value as high as 312.87 and an effect size eta squared of 0.779, confirming that the characteristics of multi-source data can effectively distinguish different skill levels.

B. EVALUATION OF THE EFFECTIVENESS OF THE TRAINING PROGRAM

Download the source code. 180 teacher trainees were randomly split into training set (126) and test set (54) (7:3 ratio). The data of training set was used to train the K-means and SVM classifier, and the data of test set was used to check the generalization ability of the model. This approach ensures that the model identifies skill levels consistent with professional pedagogical standards rather than relying solely on unsupervised data patterns, thereby avoiding circularity in the training and evaluation process.

Confusion matrix was calculated and the number of true positive, false positive, true negative, false negative were counted. Then, we could get the accuracy, precision, recall and F1 score of the four performance metrics. Finally, the five-fold cross validation was used to repeat the above experiment 100 times, and then we could get the mean and standard deviation of the performance.

Compared the classification performance of SVM with the three state-of-the-art algorithms: random forest, gradient boosting tree, multilayer perceptron. In addition, we also recorded the training time and prediction time of these algorithms to evaluate the computational efficiency. The data are shown in Table 3.

TABLE 3. Comparison of Classification of Skill Diagnostic Models

Model Algorithm	Accuracy (%)	Accuracy (%)	Recall rate (%)	F1 score (%)	Training time (s)
SVM (RBF kernel)	87.6±2.3	86.9±2.8	87.2±2.5	87.0±2.4	3.8±0.6
Random Forest	82.4±3.1	81.7±3.5	82.1±3.2	81.9±3.3	5.2±0.8
Gradient boosting tree	84.1±2.7	83.5±3.0	83.8±2.9	83.6±2.8	7.6±1.2
Multilayer perceptron	80.3±3.8	79.6±4.2	80.0±4.0	79.8±4.1	12.4±2.1
Baseline (Single Video Feature)	73.5±4.5	72.8±4.9	73.1±4.7	72.9±4.8	2.1±0.4
Baseline (single eye movement feature)	71.2±5.1	70.4±5.6	70.8±5.3	70.6±5.4	1.9±0.3

The SVM model achieved an accuracy of 87.6% (95% CI: [85.3%, 89.9%]) on the test set, outperforming the other three algorithms by 5.2 to 7.3 percentage points. Paired t-tests conducted across the five-fold cross-validation folds confirmed that SVM's accuracy was significantly higher than that of Random Forest ($p< 0.05$) and Multilayer Perceptron ($p< 0.01$). A Cohen's Kappa coefficient of 0.82 indicates that the classification results are almost perfectly consistent with expert assessments. The balance between precision and recall kept the F1 score at 87.0%, a 14.1 percentage point improvement in accuracy compared to the baseline model using only a single video feature, confirming the effectiveness of multi-source data fusion. The SVM training time of only 3.8 seconds is significantly shorter than gradient boosting trees and multilayer perceptrons, meeting the efficiency requirements for real-time diagnosis. The standard deviation of the five-fold crossvalidation is controlled within 2.3%, indicating good model stability and resistance to sample fluctuations.

C. VALIDATION OF THE IMPROVEMENT PATH

To evaluate the effectiveness of the three-stage training program, a pre-test and post-test design was implemented for the 180 trainees. Following the 8-week intervention, participants demonstrated a significant improvement in their Overall Skills Score (Pre-test: 67.1 $\pm$ 9.4; Post-test: 79.8 $\pm$ 6.2; $t(179) = 14.52$, $p< 0.001$). Notably, the low-level group showed the largest gain, with a 34% increase in 'Percentage of valid observations.' Furthermore, a control group of 40 trainees who received traditional internship supervision without the multi-source feedback showed significantly lower improvement ($p< 0.05$), confirming that the data-driven personalized intervention effectively bridges the identified skill gaps.

V. CONCLUSION

Multi-source data fusion technology can solve the single-dimensional assessment limitation by combining objective indicators at three levels: behavior, cognition, and physiology to build a precise portrait of early childhood teacher trainees' outdoor play guidance skills. Machine learning method has good classification performance and generalization ability in the automatic skill level identification, and provides a feasible technical solution for large-scale diagnostic assessment of teacher trainees' ability. Prelim-

inary analysis of the pilot implementation suggests that the three-tiered progressive training program has the potential to enhance skill acquisition by integrating virtual simulation training and real practice; future longitudinal studies will incorporate pre- and post-test experimental designs to quantify these improvements; the complementability between the controllability of virtual environment and the authenticity of on-site guidance enhance the skill transfer. Eye-tracking explores the visual attention allocation in different situations from a cognitive neuroscience perspective to understand the expert-novice difference, and meanwhile, physiological signal monitoring provides an index to reflect the implicit difference in emotion regulation ability.

AUTHOR CONTRIBUTIONS

Laling Dong designed, collected and analyzed the data, and drafted the manuscript. Laling Dong conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Laling Dong participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDINGS

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of Jinan Preschool Education College, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent for

REFERENCES

[1] F. Xu, X. Guo, J. Pan, et al., "An Innovative Guidance for Outdoor Fitness Equipment Design," Lecture Notes in Computer Science, 2022, doi: 10.1007/978-3-031-06050-2_31.

[2] Z. Cai and G. Xing, "Balancing Freedom and Guidance: Kindergarten Teachers' Strategies for Outdoor Play (in the Post-COVID Era)," Proceedings of the International Conference of the Learning Sciences, pp. 1754-1757, 2024, doi: 10.22318/icls2024.754585.

[3] A. Rovira, G. Schieck A F, P. Blume, et al., "Guidance and surroundings awareness in outdoor handheld augmented reality," PLoS ONE, vol. 15, no. 3, Art. no. e0230518, 2020, doi: 10.1371/journal.pone.0230518.

[4] G. Zhang, J. Sun, J. Gong, et al., "Exploration of visual variable guidance in outdoor augmented reality geovisualization," International Journal of Digital Earth, vol. 16, no. 2, pp. 19, 2023, doi: 10.1080/17538947.2023.2259874.

[5] P. Levinger, M. Panisset, H. Parker, et al., "Guidance about age-friendly outdoor exercise equipment and associated strategies to maximise usability for older people," Health promotion journal of Australia: official journal of Australian Association of Health Promotion Professionals, 2020, doi: 10.1002/hpja.367.

[6] Ponomareva L I , Krezhevskikh O , Kolmogorova I V ,et al.Preparation of teachers for the organization of early career guidance in a preschool educational organization[J].Perspectives of Science and Education, 2023, doi: 10.32744/pse.2023.2.40.

[7] Y. Hei, S. Suhaily S, J. Wu, et al., "Proposition on the Redesigning Outdoor Guidance Systems on Campus: A Case Study of Southwest Minzu University Aeroport Campus in Southwest China," Knowledge-Based Engineering and Sciences, 2022, doi: 10.51526/kbes.2022.3.2.52-65.

[8] Zhe Naiping, "Releasing Nature and Enjoying Games: An Exploration of Guidance Strategies for Children's Outdoor Independent Play," Journal of Learning, vol. (23), pp. 2, 2021, doi: 10.16657/j.cnki.issn1673-9132.2021.23.089.

[9] D. Craig, N. A. Trina, M. Monsur, et al., "Effective nature-based outdoor play and learning environments for below-3 children: a literature-based summary," International Journal of Environmental Research and Public Health, vol. 21, no. 9, pp. 1247, 2024, doi: 10.3390/ijerph21091247.

[10] R. Parker and S. Al-Maiyah, "Developing an integrated approach to the evaluation of outdoor play settings: Rethinking the position of play value," Children's Geographies, vol. 20, no. 1, pp. 1-23, 2022, doi: 10.1080/14733285.2021.1912294.