
5. CHAPTER 5: DISCUSSION, LIMITATIONS, AND FUTURE WORK

5.1. Discussion of Findings

5.1.1. Hypothesis H1: Audio-Temporal Alignment

RQ1 asked how audio-based cross-correlation contributes to the alignment of student and reference dance videos. The implementation results show that temporal alignment is necessary in the proposed system because student recordings do not naturally begin at exactly the same moment as the reference video. Without an alignment stage, the pose comparison process would compare student movement frames against mismatched reference frames, producing unreliable similarity scores even when the student performs the correct movement.

The audio cross-correlation stage addresses this problem by estimating the temporal offset between the student and reference audio signals before pose-based comparison is performed. This allows the system to anchor the student video to the musical structure of the reference performance. After this global synchronization step, FastDTW is applied to absorb smaller local variations in movement timing.

Therefore, H1 is supported from an implementation perspective. The system successfully integrated audio-based temporal alignment into the scoring pipeline, and all 45 submitted videos were processed without extraction or alignment failure. This confirms that the combination of audio cross-correlation and FastDTW is a practical approach for handling start-time differences and minor tempo variation in video-based dance assessment.

5.1.2. Hypothesis H2: Semantic Feedback Quality

RQ2 asked whether semantic feedback provides more useful correction guidance than numerical scores alone. The system addresses this question by converting kinematic deviations into natural-language corrective feedback. Instead of presenting only numerical values such as angular deviation or similarity distance, the system generates body-part-specific instructions that are closer to the language used by dance instructors.

The semantic feedback module was successfully demonstrated across all 45 recordings. For each evaluated video, the system generated feedback based on the detected movement deviations and mapped those deviations into corrective messages. This supports the feasibility of translating pose-based measurements into understandable instructional feedback for beginner learners.

However, H2 should be interpreted cautiously. This study did not conduct a controlled experiment comparing learning improvement between users receiving numeric-only feedback and users receiving semantic feedback. Therefore, H2 is qualitatively supported in terms of system functionality and interpretability, but not fully validated as a learning-outcome claim. A longitudinal user study is required to determine whether semantic feedback measurably improves student correction and skill development over time.

5.1.3. Hypothesis H3: System-Expert Correlation

RQ3 asked to what extent the system-generated scores align with expert dance evaluation. Hypothesis H3 expected a high correlation of $r \geq 0.70$ between system-generated scores and expert scores. The evaluation results did not support this hypothesis. Across all 45 videos, the overall Pearson correlation was $r = 0.25$ ($p = 0.092$), with a Mean Absolute Error (MAE) of 17.27 points and a Root Mean Square Error (RMSE) of 20.95 points.

Although the correlation was positive, it was weak and did not reach the predefined threshold of $r \geq 0.70$. Therefore, the current scoring model cannot be considered strongly aligned with expert judgment. A systematic upward bias of +8.93 points was also observed, indicating that the system generally assigned higher scores than the expert evaluator. This suggests that the current scoring thresholds are more lenient than the expert's assessment standard.

The per-movement results show that system performance varied substantially by movement type. Bouncing produced the strongest alignment with expert evaluation, with a moderate positive correlation of $r = 0.54$ ($p = 0.037$). However, it also had a high positive bias of +17.21 points, meaning that the system ranked Bouncing performances more consistently than the other movements but still overestimated the absolute score

values.

Stepping produced the lowest MAE at 10.96 points, but its correlation was weak and statistically non-significant ($r = 0.11$, $p = 0.707$). This indicates that although the system's numerical errors were smaller for Stepping, it did not strongly reproduce the expert's ranking of student performances.

Sliding was the weakest movement-specific component. It produced a weak negative correlation ($r = -0.13$, $p = 0.651$) and the highest MAE at 23.61 points. This suggests that the current Sliding scoring features do not sufficiently capture the qualities emphasized by the expert evaluator, such as smooth lateral travel, weight transfer, hip stability, and continuous glide control.

Overall, H3 is not supported. The system functions as a working prototype for pose-based dance self-assessment, but the scoring model requires further calibration and feature refinement before it can be considered strongly aligned with expert evaluation.

5.1.4. The Wiraga-Wirama Framework as Computational Rubric

A distinctive contribution of this research is the deliberate grounding of its computer vision pipeline in the Wiraga-Wirama framework drawn from Indonesian classical dance pedagogy (Robby, 2025). Although these concepts originate in a cultural tradition far removed from hip-hop, their underlying constructs (bodily correctness and rhythmic synchronization) are universal to any choreographed movement evaluation. By mapping Wiraga onto spatial accuracy (per-frame Euclidean distance across 13 anatomical landmarks) and Wirama onto temporal synchronization (audio cross-correlation offset and FastDTW alignment quality), the system transforms abstract pedagogical ideals into concrete, computable signal-processing operations.

This bridge between traditional dance philosophy and modern artificial intelligence is not merely rhetorical. It provided a principled basis for the 50/50 weighting of Wiraga and Wirama in the composite score, endorsed by Coach Ambrosius Robby, and it ensured that every design decision in the pipeline (from the choice of time-series metrics per movement type to the four-band rubric) was anchored

in established evaluation practice rather than arbitrary engineering choices. Future extensions of the system, including the potential quantification of Wirasa (expressive feeling), will benefit from the same philosophy: begin with what dance teachers already know, then ask which signals best capture it computationally.

5.2. System Limitations

5.2.1. Single Reference Per Movement

The system currently relies on a single expert-validated reference video per movement type. While this reference was recorded and approved by Coach Ambrosius Robby, a practitioner with twenty years of professional experience, it represents one interpretation of each movement performed by one body at one tempo. Differences in teaching style across schools, natural variation in body proportions (limb length ratios, torso height), and differing tempo preferences across instructors are not captured by a single reference. Students whose anatomy or training background differs substantially from the reference performer may receive systematically lower Wiraga scores even when their form is technically correct by other standards. Expanding the reference library to cover multiple expert performers is therefore a prerequisite for generalizing the system beyond its current evaluation context.

5.2.2. Absence of Movement Type Classifier

The system trusts the user's declared movement type when selecting the appropriate scoring pipeline. No automatic classifier verifies that the uploaded video matches the declared category. Empirical testing confirmed the severity of this gap: a Stepping video submitted under the Bouncing category returned a Wiraga score of 77 out of 100, a score that would mislead both the student and any instructor reviewing it. The mismatch arises because the Bouncing pipeline applies knee-bend turning-point analysis to footage that contains foot-strike patterns instead, and the partial signal overlap produces plausible but incorrect results. A machine learning classifier trained on the three movement types would close this vulnerability; its development is proposed in Section 5.3.1.

5.2.3. Wirasa Dimension Not Implemented

The traditional three-pillar framework from which this research draws its evaluation philosophy (Wiraga, Wirama, and Wirasa) is only two-thirds implemented. Wirasa, which describes the expressive and emotional quality of a performance, cannot be reliably quantified from 2D skeletal pose landmarks alone. Emotional expression in dance is communicated through facial nuance, gaze direction, breath timing, muscle tension, and stylistic intent: signals that fall outside the information space of MediaPipe BlazePose's 33-landmark skeleton. The system therefore scores only two of the three traditional quality dimensions, and students who have mastered expressive performance will not see that mastery reflected in their scores. As discussed in Section 5.3.4, energy dynamics such as acceleration contrast and movement amplitude variance may serve as a partial computational proxy for Wirasa in future iterations.

5.2.4. Camera Angle Dependency

The entire pipeline is optimized for frontal camera setups in which the student faces the lens directly. MediaPipe BlazePose's landmark detection confidence degrades substantially when the body is viewed from oblique or rear angles, particularly for joints that become self-occluded: the far hip, the trailing knee, or the arms crossing the torso midline. In home practice environments, camera placement is rarely controlled, and students may position their phones at unconventional angles to capture more of the room. The current visibility gating mechanism ($v < 0.60$ threshold with temporal fallback) mitigates transient occlusion events but cannot compensate for systematic angle-induced uncertainty across an entire recording. Multi-camera setups or view synthesis would be required to fully resolve this limitation.

5.2.5. Fixed Scoring Thresholds

Key scoring thresholds (including the knee bend target of 90° for Bouncing evaluation, the timing tolerance of 100 milliseconds for Wirama scoring, and the 12° base joint-deviation threshold for semantic feedback triggering) are calibrated to the standards endorsed by Coach Ambrosius Robby for intermediate-level dancers. These values are not adaptive. A beginner for whom a 60° knee bend represents genuine progress will receive the same score penalty as a more advanced student executing the same angle, because the system has no mechanism for tracking or adjusting to

individual skill trajectories. Adaptive thresholds that soften as a student improves, and tighten again as new movement types are introduced, would substantially increase the system's pedagogical value for long-term users.

5.2.6. Evaluation Dataset Size

The evaluation was conducted with five student participants producing 45 videos in total, consisting of three movements and three attempts per student. While this dataset was sufficient to demonstrate system functionality and conduct preliminary correlation analysis, it remains too small for strong statistical generalization.

The skill-level grouping also has an important limitation. The participant distribution consisted of two Basic-level students, one Intermediate-level student, and two Advanced-level students. Because the Intermediate category contained only one participant, skill-level comparisons should be interpreted only as descriptive context rather than inferential evidence. The main evaluation focus of this study was the relationship between system-generated scores and expert scores across all 45 videos, not the statistical comparison of proficiency groups.

A larger and more balanced dataset would be necessary to evaluate whether the system can reliably distinguish learner skill levels. Future studies should recruit more participants across Basic, Intermediate, and Advanced categories, ideally with a balanced number of students in each group.

5.3. Future Work

5.3.1. Movement Type Auto-Detection

The most immediately actionable extension is the development of an automatic movement type classifier. A lightweight model (for example, a recurrent neural network or a temporal convolutional network trained on the joint angle time-series of the three current movement categories) could verify at upload time that the submitted video corresponds to the declared movement type. Beyond preventing the scoring errors documented in Section 5.2.2, such a classifier would lay the groundwork for unsupervised movement labeling, enabling the system to eventually score performances without requiring the student to select a category at all.

5.3.2. Adaptive Skill-Level Thresholds

The fixed thresholds described in Section 5.2.5 should be replaced with dynamic, user-history-aware equivalents. A proficiency model that tracks a student's rolling average score across sessions could progressively tighten the knee-bend target, timing tolerance, and feedback trigger threshold as performance improves, maintaining appropriate challenge without demotivating beginners. Such adaptive thresholds mirror the behavior of skilled human teachers, who naturally adjust their standards to the student standing in front of them, a pedagogical pattern that the current system's fixed calibration cannot replicate.

5.3.3. Multi-Reference Dataset

Building a curated library of ten to twenty expert-validated reference videos per movement type would substantially improve scoring robustness. Rather than comparing a student against a single reference, an ensemble comparison approach could compute the student's score relative to each reference in the library and report the median, reducing the sensitivity of the Wiraga score to idiosyncrasies of any one expert performance. This approach would also naturally accommodate variation in teaching style and body type, addressing the generalization concern raised in Section 5.2.1.

5.3.4. Wirasa Proxy Measurement

Although 2D pose landmarks cannot capture emotional expression directly, certain kinematic signatures may correlate sufficiently with what expert observers recognize as expressiveness to serve as useful proxies. Acceleration contrast (the ratio between sharp, accented movements and sustained, flowing ones) captures stylistic dynamism. Movement amplitude variance across repeated cycles reflects the consistency and commitment of expressive choices. A dedicated study, in which expert coaches rate expressiveness on a structured scale while the system logs these proxy metrics, would determine whether a meaningful computational approximation of Wirasa is achievable within the current sensing modality.

5.3.5. Real-Time Browser-Side Processing

The current pipeline requires students to upload a completed recording and wait

for server-side processing before receiving feedback. Migrating the pose extraction layer to MediaPipe.js (the JavaScript port of the BlazePose pipeline) would enable landmark detection to run entirely in the student's browser, eliminating upload latency and enabling real-time feedback during practice. A student could observe their current Wiraga score updating live on screen as they perform, closing the feedback loop from minutes to milliseconds. This architectural change would also reduce server infrastructure costs and improve the system's accessibility in low-bandwidth environments.

5.3.6. Extended Movement Vocabulary

The three movements currently supported (Bouncing, Stepping, and Sliding) represent foundational hip-hop elements but do not exhaust the movement vocabulary that students and teachers encounter in practice. Extending the system to cover additional categories such as popping, locking, and top rock would require developing new movement-specific Wiraga scoring pipelines in collaboration with expert coaches, collecting reference recordings, and validating scoring thresholds for each new category. Each addition follows the same development methodology established in this thesis, making systematic expansion a tractable engineering task once the evaluation infrastructure is in place.

5.3.7. Longitudinal Improvement Study

Perhaps the most important open question left by this research is whether the system actually helps students improve. The 45-video dataset captures a single cross-sectional evaluation; it says nothing about whether students who use the system over multiple weeks develop better technique than those who practice without it. A longitudinal study tracking ten to fifteen students across six to eight weeks (with pre-tests, mid-point assessments, and post-tests scored by both the system and an independent expert judge) would provide the evidence base needed to validate the system's pedagogical effectiveness and, by extension, to establish whether H2's claim about the value of semantic feedback holds in real learning conditions.

5.3.8. Scalability to Larger Student and Instructor Participation

The current system was implemented and evaluated as a prototype using five student participants and one expert-validated reference video for each movement. This scope is sufficient to demonstrate the end-to-end workflow, including video upload, pose extraction, temporal alignment, scoring, semantic feedback, and usability evaluation. However, the current implementation should not be interpreted as evidence that the system has already been validated for large-scale use across many students, instructors, body types, or dance schools.

As the number of student users increases, the system would require a more scalable deployment environment than the current prototype configuration. A larger deployment should use stronger server resources, structured storage for uploaded videos and generated pose sequences, efficient database management, and queue-based processing for multiple video submissions. The current SQLite-based implementation is appropriate for prototype-scale evaluation but would need to be replaced or extended with a production-grade database and storage strategy for wider use.

As the number of instructors and reference performers increases, the system would also require a multi-reference design. A single reference represents one body proportion, movement style, tempo preference, and instructional interpretation. Future work should therefore create a reference library containing multiple expert performers, multiple skill levels, and multiple acceptable versions of each movement. Expert ratings from more than one coach should also be collected to establish inter-rater reliability and reduce dependence on one evaluator's scoring tendency.

A larger student-and-instructor dataset would support more advanced improvements, including adaptive scoring thresholds, personalized reference selection, automatic movement classification, and machine-learning models trained on broader variation. However, these improvements should be introduced only after collecting sufficiently diverse data and validating the resulting system against multiple expert evaluators.

5.4. Conclusion

This thesis designed, implemented, and evaluated a video-based AI-assisted motion tracking system for foundational hip-hop dance training. The system

operationalizes the Wiraga and Wirama evaluation framework as a two-component computational rubric applied to three beginner-level movements: Bouncing, Stepping, and Sliding. Kinematic data is extracted from standard video using MediaPipe BlazePose, processed through visibility filtering and smoothing, and aligned to a reference performance using audio cross-correlation followed by FastDTW. Movement-specific scoring algorithms are then combined with a Wirama timing score to produce a composite assessment. A semantic feedback layer translates detected movement deviations into natural-language corrective instructions, allowing students to interpret the system output without requiring technical knowledge of pose estimation or kinematic analysis.

The evaluation involved five student participants across three skill levels, producing 45 videos assessed independently by an expert dance coach. The system processed all 45 videos successfully, demonstrating that the implemented pipeline is functional as a lightweight web-based self-assessment prototype. The usability evaluation also produced a mean SUS score of 69.00, indicating acceptable usability slightly above the commonly used average benchmark.

However, the system-expert correlation results shows that the current scoring model is not yet strongly aligned with professional evaluation. The overall Pearson correlation between system-generated scores and expert scores was $r = 0.25$ ($p = 0.092$), which did not meet the H3 threshold of $r \geq 0.70$. The system also showed a positive bias of +8.93 points, indicating that it generally scored more leniently than the expert evaluator. Among the three movements, Bouncing showed the strongest alignment with expert judgment, Stepping produced the lowest absolute error, and Sliding remained the weakest component due to its weak negative correlation and highest MAE.

These findings indicate that the system should not be interpreted as a replacement for professional dance instruction. Instead, it is better positioned as an educational support tool that provides structured self-assessment and feedback for beginner learners. The main contribution of this thesis is the implementation and evaluation of an integrated framework combining pose estimation, temporal alignment, movement-specific scoring, and semantic feedback in a lightweight web-based dance training system.

Future development should focus on recalibrating scoring thresholds, improving Sliding-specific feature extraction, expanding the reference dataset, adding movement type auto-detection, and conducting a larger longitudinal study to evaluate whether semantic feedback contributes to measurable student improvement. The results of this study show both the feasibility and the limitations of pose-based dance assessment, providing a foundation for future work in AI-assisted movement learning systems.



REFERENCES

- Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, 100215. <https://doi.org/10.1016/j.caeai.2024.100215>
- Art Gharana Academy. (2025). *Modern Hip Hop Pedagogy and Movement Classification for Digital Learning*.
- Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). BlazePose: On-device Real-time Body Pose Tracking. *ArXiv Preprint ArXiv:2006.10204*. <https://doi.org/10.48550/arXiv.2006.10204>
- Fink, J. (2025). *Foundational Hip Hop: A Guide for Educators and Students*. Human Kinetics.
- Kang, J., Kang, C., Yoon, J., Ji, H., Li, T., Moon, H., Ko, M., & Han, J. (2023). Dancing on the inside: A qualitative study on online dance learning with teacher-AI cooperation. *Education and Information Technologies*, 28(9), 12111–12141. <https://doi.org/10.1007/s10639-023-11649-0>
- Kim, J.-W., Choi, J.-Y., Ha, E.-J., & Choi, J.-H. (2023). Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model. *Applied Sciences*, 13(4), 2700. <https://doi.org/10.3390/app13042700>
- Lee, S., Kim, J., & Park, Y. (2023). A Comparative Study of Human Motion Recognition Based on Pose Estimation. *IEEE Access*, 11. <https://doi.org/10.1109/ACCESS.2023.3278914>
- Lee, S. ; W. R. (2023). Angular Kinematics for View-Invariant Human Pose Assessment in Dance. *IEEE Transactions on Multimedia*, 25(412–425).
- Mroz, S., Baddour, N., McGuirk, C., Juneau, P., Tu, A., Cheung, K., & Lemaire, E. (2021). Comparing the Quality of Human Pose Estimation with BlazePose or OpenPose. *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*, 1–4. <https://doi.org/10.1109/BioSMART54244.2021.9677850>
- Niu, C. (2024). The application of improved DTW algorithm in sports posture recognition. *Systems and Soft Computing*, 6, 200163. <https://doi.org/10.1016/j.sasc.2024.200163>
- Niu, C., Liu, J., & Zhang, W. (2024). Application of Improved Dynamic Time Warping Algorithm in Sports Posture Analysis. *Journal of Intelligent Systems*. <https://doi.org/10.1016/j.jintsys.2024.02.009>

- Wang, L., Zhao, H., & Zhou, X. (2021). A Deep Learning-Based Motion Analysis for Dance Performance Evaluation. *Sensors*, 21(15), 5145. <https://doi.org/10.3390/s21155145>
- Yu, X., Lee, B., & Sedlmair, M. (2024). Design Space of Visual Feedforward And Corrective Feedback in XR-Based Motion Guidance Systems. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3613904.3642143>
- Robby, A. (2025). Expert validation of Wiraga-Wirama scoring framework for basic hip-hop dance movements [Unpublished expert consultation]. Independent Dance Coach and Practitioner, 20 years professional experience.



SAMUEL KEVIN LALUYAN

Tangerang, Banten, 15151

+62 81314512259 | samueltjla90@gmail.com | [linkedin.com/in/samuel-kevin-laluyan](https://www.linkedin.com/in/samuel-kevin-laluyan)

PROFESSIONAL SUMMARY

Highly motivated final-year IT student with a strong interest in project coordination and teamwork. Trained in a technical background and structured approach for problem-solving, building positive relationships, and communicate effectively across teams. Eager to contribute to project success by supporting planning, execution, and delivering efficient and well-organized outcomes.

WORK EXPERIENCE

PT Anugerah Lelang Indonesia *Full-Stack Mobile App Developer (Intern)* | 2024 – 2025

- Collaborated with a development team to design and launch the smartBiD mobile application, ensuring deadlines were met and user feedback was implemented.
- Was given an opportunity to be a **Project Manager** in Mobile Dev department, familiar using project management software such as **Trello** and **Jira**.

SKILLS & CERTIFICATIONS

- **Programming:** C++, Java, Python, Dart, Go, PHP, Laravel, HTML, CSS.
- **Project & Collaboration Tools:** Jira, Trello (task tracking, workflow management).
- **Technical Skills:** UI/UX Design, Front-End, Back-End, Middleware, Network Security
- **Certifications:**
 - Network Defense Essentials (NDE)
 - Ethical Hacking Essentials (EHE)
- **Languages:** Bahasa Indonesia (Fluent), **English (Professional Working Proficiency)**, German (Basic)
- **Soft Skills:** Teamwork, Leadership, Communication, Management Skills, Problem-solving, Critical Thinking, Creativity.

ORGANIZATIONAL EXPERIENCE

General Elections Commission (KPPS) - Member | 2024

- Managed high-volume crowds in a fast-paced environment and assisted individuals calmly while resolving issues under pressure.

Swiss German University **Modern Dance UKM - Head** | 2022 – 2023

- Led a team, coordinated schedules, and maintained energy and professionalism during club activities.

PENABUR Bintaro Jaya High School **Student Council – Treasurer (Board of Directors)** | 2021 – 2022

- Entrusted with handling large sums of cash and managing financial reports for entire student council organization.

EDUCATION

Swiss German University Bachelor of Information Technology | 2022 – Present

PENABUR Bintaro Jaya High School Science Program | 2019 – 2022