# Automated Stress Vulnerability Assessment from Person-in-the-Rain Drawing Test Using Multimodal Deep Learning

Syed Ali Naqi Raza[1], *Senghak Lim[1],* Chomyong Kim[2], Yunyoung Nam[3]
[1]*Department of ICT Convergence, Soonchunhyang University, Asan, 31538, Republic of Korea*
[2]ICT Convergence Research Centre, Soonchunhyang University, Asan, South Korea
[3] Department of Computer Science and Engineering, Soonchunhyang University, Asan, South Korea
*Contact: ynam@sch.ac.kr

*Abstract*— **In clinical psychology, projective drawing tests like the Person-in-the-Rain (PITR) are commonly used to examine stress perception and coping actions. Nevertheless, traditional scoring depends largely on expert judgment, which can bring in subjectivity and restrict scalability. This research suggests a multimodal deep learning framework that combines visual data from drawings with corresponding text descriptions to facilitate automated psychological stress assessment.Participants are classified into two psychological states: Resilient to Stress, characterized by coping features surpassing stress indicators; Vulnerable to Stress, where stress-related features predominated. The suggested architecture uses a deep neural image encoder to extract structural and semantic features from PITR drawings, together with a transformer-based text encoder for capturing contextual and descriptive information. A joint feature space for classification is created by integrating the learned representations from both modalities. It has been shown through experiments that the multimodal techniques increases classification performance in comparison to single-modality baselines, demonstrating the complementary nature of visual and textual cues in stress evaluation.**

## I. Introduction & Related Work

Emotional control, behavioral adaptation, and long-term mental health effects are all significantly influenced by psychological stress, specifically in children and adolescents. It is essential for preventive intervention and targeted psychological support to identify stress vulnerability and coping capacity early on. Among projective assessment techniques, the Draw-A-Person-in-the-Rain (DAPR) or Person-in-the-Rain (PITR) test is commonly used in clinical psychology to investigate perceived stress and coping mechanisms through symbolic drawing elements [1][2]. This evaluation uses environmental stress indicators like rain amount, angle of rain, shape of rain, presence of puddles and clouds to reflect perceived stress, on the other hand protective elements like umbrella or shelters symbolize coping resources.

Despite it is clinically relevant, traditional scoring of PITR depends highly on subjective interpretation of clinical experts. Trained psychologists are required for manual scoring, which can lead to inter-rater variability. Reproducibility and scalability is affected by this limitation in large scale screening contexts. The ongoing evolution of healthcare and behavioral analysis by artificial intelligence, automated psychological drawing assessment creates a promising avenue for research.

Current developments in multimodal learning and computer vision have enabled automated analysis of drawings and sketches. Convolutional Neural Networks and Vision Transformers have demonstrated strong efficacy in extracting high-level visual features from images [4][5]. While, language models built on transformers like BERT and DistilBERT have attained cutting-edge performance in semantic representation learning [6]. Multimodal learning is further facilitated by joint image-text models like CLIP [7].

Automated systems are making significant contribution in the fields of art therapy and projective assessments. Kim et al. Proposed an explainable AI-based expert support system for art therapy [8]. CheckDAPR, a system for Draw A Person In Rain test assessments based on MLLM, was introduced by Yang et al. [9]. Posada Barrera and Pena Diaz developed a deep learning system for the generation of structured information from DAPR drawings. Furthermore, psychological screening has utilized AI methods, using hand drawing tests like clock drawing for dementia screening and sketch based cognitive assessment [11][12].

These researches mainly emphasizes on structured feature extraction or explainability, rather than stress vulnerability. Moreover, limited research integrated both the visual drawing features and contextual descriptions for psychological state profiling. Clinical interpretation is essentially multimodal, psychologists consider both structural drawing patterns and narrative descriptions when analyzing stress and coping dynamics.

In this study we propose a multimodal deep learning framework that integrates image and text representations to automate the stress vulnerability assessment from PITR hand drawings. A coping capacity score calculated that balances between stress indicators and coping resources as depicted in drawings. Participants are categorized into resilient, and vulnerable profiles based on this structures representation.

Our contributions are summarized as follows:

1. A multimodal deep learning framework for stress vulnerability classification from PITR drawings.
2. A comparative evaluation of unimodal and multimodal architectures.
3. An objective and scalable approach for AI-assisted psychological screening.

The remainder of this paper is organized as follows: Section 2 presents the proposed method, including subject selection, data acquisition, and details of VR tasks. Section 3 presents the results and discusses the findings in detail. Finally, Section 4 summarizes the research by providing a conclusion and final remarks on the work conducted, along with suggestions for future work.

## II. Methodology

### *A. Participants Selection:*

We recruited 196 participants. The participants were able to follow simple drawing instructions, without any diagnosed visual problem. Informed consent was obtained from participants before participation in experiment.

### *B. Person-in-the-Rain (PITR) Test Procedure:*

The Person-in-the-Rain (PITR), known as Draw-A-Person-in-the-Rain (DAPR), is a projective drawing assessment designed to assess perceived stress and coping capacity [1], [2]. The test operates on the premise that environmental elements shown in the drawing represent perceptions related to stress, while protective elements denote coping resources. Clinical interpretation of PITR drawings focuses on two primary dimensions:

Stress Indicators (RS):
Rain amount and intensity, shape of rain, wind direction, rain touching person, rain touching clouds, number of clouds, and area, number and size of lightning bolts, lightning person distance, number and width of puddles, puddle person distance. Greater intensity, density, or environmental dominance is interpreted as higher perceived stress exposure.

Coping Resources (CR)
Number of direct protective objects (umbrella, raincoat, boots, hat, etc. used to avoid rain), Number of indirect protective objects (house, tree, bag, car, etc. excluding item 1, that can block rain), total number of protective objects, adequacy of direct protection (ability to protect the person), adequacy of indirect protection (other protective means), facial coverage (degree to which the face is covered), facial expression (overall expression), size of the person, position of the person, completeness of the person, stroke quality (line quality). More prominent protective features indicate stronger perceived coping capacity.

TABLE I

STRESS INDICAORS

| No. | Item | Detailed Critera |
|---|---|---|
| 1 | Rain Amount | Number of raindrops |
| 2 | Rain Intensity | Length of rain lines |
| 3 | Shape of Rain | — |
| 4 | Wind Strength | Angle of rain |
| 5 | Rain touching person | Degree of contact |
| 6 | Rain touching clouds | Degree of contact |
| 7 | Number of clouds | — |
| 8 | Cloud area | Cloud width |
| 9 | Number of lightning bolts | — |
| 10 | Power of lightning | Size of lightning |
| 11 | Lightning-person relation | Distance |
| 12 | Number of puddles | — |
| 13 | Puddle area | Width of puddles |
| 14 | Puddle-person relation | Distance |

TABLE II

COPING INDICAORS

| No. | Item |
|---|---|
| 1 | Number of direct protective objects (umbrella, rainc oat, boots, hat, etc. used to avoid rain) |
| 2 | Number of indirect protective objects (house, tree, b ag, car, etc. excluding item 1, that can block rain) |
| 3 | Total number of protective objects |
| 4 | Adequacy of direct protection (ability to protect the person) |
| 5 | Adequacy of indirect protection (other protective m eans) |
| 6 | Facial coverage (degree to which the face is covere d) |
| 7 | Facial expression (overall expression) |
| 8 | Size of the person |
| 9 | Position of the Person |
| 10 | Completeness of the person |
| 11 | Stroke quality (line quality) |



*Figure 1 Sample of Person in Ratin Test*

*C. MultiModality:*

The proposed multimodal framework seeks to analyze together the structural features of PITR drawings and their related semantic descriptions. Scores of Stress Indicators and Coping Resource are converted into textual description for text modality along with image modality for better contextual awareness. Figure 1 shows the flow of study.

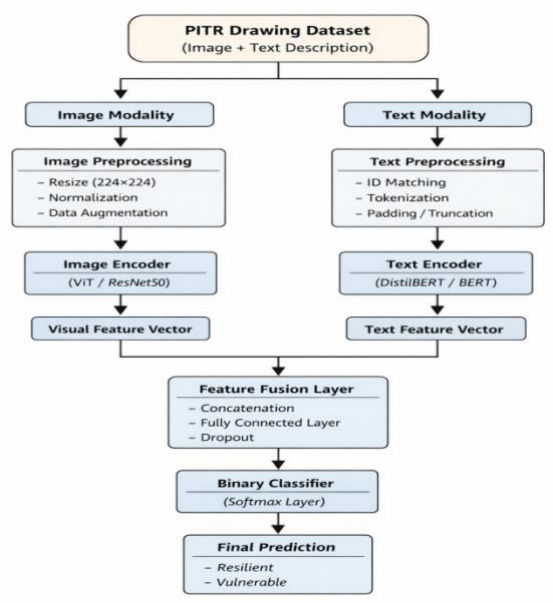



*Figure 2 Overall flow of study*

*Visual representation Learning:*

We utilized a pretrained deep convolutional and transformer-based image encoders that extracts discriminative structural features from PITR drawings. The classification heads of these pretrained models were removed to extract only feature embeddings.

*ResNet-50 Backbone*

Resnet-50, a deep convolutional neural network, comprises of residual blocks that facilitate stable gradient propogation in deep architectures. When given an input image $I \in \mathbb{R}^{224x\ 224\ x\ 3}$, the network generates a high level feature embedding:

$$v\ =\ fResNet\ (I), v\ \in \mathbb{R}^{d_v}$$

Where dv denotes the dimensionality of the global pooled feature vector.

*Vision Transformer*

Vision Transfomers divides the image into fixed size small patches and models long range spatial dependencies using self attention. Unlike CNNs, ViT directly captures global structural relationships for the drawing.

$$I \rightarrow \{p_1, p_2, \ldots . p_n\}$$

These patches are embedded and went through transformer layers. The final CLS token mbedding is provides as:

$$v\ =\ f_{VIT}(I)$$

*Textual representation Learning:*

The textual descriptions attached with each drawing were encoded using DistilBERT. It's a lightweight transformer model which is pretrained on large-scale corpora. Given a tokenized sequence T, the encoder produces a contextual embeddings:

$$t\ =\ f_{DistilBERT}(T), t \in\ \mathbb{R}^{d_t}$$

*Multimodal Fusion Strategy:*

A late fusion strategy was utilized to integrate visual and textual representations. The visual and textual embeddings are concatenated as:

$$z = [v; t]\ \in\ \mathbb{R}^{d_v+d_t}$$

It allows every modality to independently learn specialized representations before integration. The fused vector z is passed to a fully connected multilayer perceptron.

$$h = \sigma(W_{1z} + b_1)$$

$$\boldsymbol{o = W_2 h\ +\ b_2}$$

Where W1, W2 are learnable weight matrices, $\sigma$ denotes ReLU activation.

The final logits $o \in \mathbb{R}^2$ are changed to class probabilities using softmax:

$$y\ =\ Softmax\ (o)$$

## III. RESULTS

### D. Image Modality Results:

Table 3 shows the performance results of image-only models for classification of PITR drawings into binary classes Resilient and Vulnerable. Two architectures are evaluated: Vision Transformer and ResNet-50.

The ViT model gained an accuracy of 0.66, having macro F1-score of 0.40 and weighted F1-score of 0.53. Since the model achieved moderate accuracy, the low macro precision shows limited ability to distinguish between classes. The confusion matrix shows that ViT correctly classified all resilient drawings, but failed to accurately identify vulnerable drawings, showing strong bias towards the majority class.

The ResNet-50 model achieved lower accuracy 0.60, as compared to ViT which achieved 0.66. However, ResNet-50 showed improved vulnerable samples detection, although some misclassifications exsisted. The confusion matrix shows the model accurately classified eight reilient samples and single vulnerable sample, while most vulnerable images are incorrectly predicted as resilient.

These results shows that image-only models captured structural drawing features, but they faced difficulty in reliably identifying indicators of vulnerability, specifically when stress related cues are context dependent.

TABLE III

IMAGE MODALITY RESULTS FOR RESILIENT AND VULNERABLE TO STRESS

| Model | | Preicision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|
| ViT | Macro Avg | 0.33 | 0.50 | 0.40 | 0.66 |
| | Weighted AVg | 0.44 | 0.66 | 0.53 | |
| ResNet-50 | Macro Avg | 0.50 | 0.50 | 0.48 | 0.60 |
| | Weighted AVg | 0.55 | 0.60 | 0.56 | |

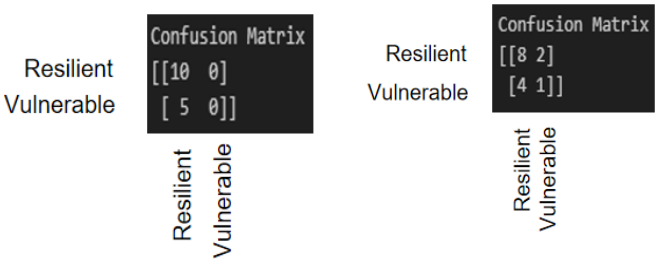



*Figure 3 (a,b) ViT and ResNet-50 Confusion Matrix*

### E. Multimodal Results:

Limitations of image-only models are addressed by using multimodal architectures by combining image features with textual descriptions. Table 4 shows the results of three multimodal models. These models include ResNet50 with DistilBERT, ViT with DistilBERT, and CLIP-based representation.

ResNet-50 and DistilBERT models achieved the highest accuracy of 0.80, macro F1-score of 0.78, and weighted F1-score of 0.80. The confusion matrix depicts balanced predictions for both classe, accurately identifying reilient and vulnerable images. This shows that integrating semantic textual information significantly improved stress classification.

On the other hand, ViT and DistilBERT model showed similar performance to image-only ViT model. It achieved 66% accuracy and a macro F1-score of 0.40. This indicates that ViT features are les compatible with simple late-fusion strategy used in the study.

The CLIP-based model, learned combined joint image-text representations. It achieved 80% accuracy and a strong macro precision of 0.88. The confusion matrix showed that CLIP correctly classified high resilient samples and also improved detection of vulnerable images compared with unimodal models.

TABLE IV

MULTI MODALITY RESULTS FOR RESILIENT AND VULNERABLE TO STRESS

| Model | | Preicision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|
| ResNet50 + DistilBERT | Macro Avg | 0.77 | 0.80 | 0.78 | 80 |
| | Weighted AVg | 0.81 | 0.80 | 0.80 | |
| ViT + DistilBERT | Macro Avg | 0.33 | 0.50 | 0.40 | 66 |
| | Weighted AVg | 0.44 | 0.66 | 0.53 | |
| CLIP model | Macro Avg | 0.88 | 0.70 | 0.72 | 80 |
| | Weighted AVg | 0.84 | 0.80 | 0.77 | |

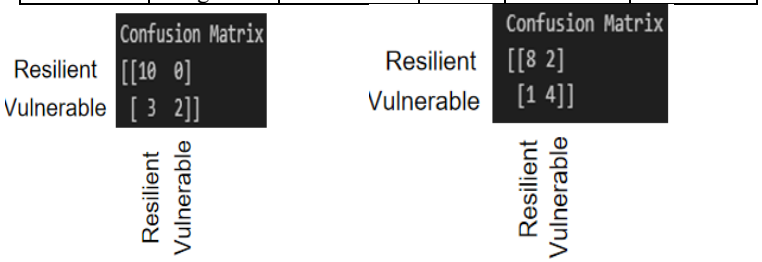



*Figure 4 (a,b) CLIP and ResNet-50 + DistilBERT Confusion Matrix*

## IV. CONCLUSION

This study discussed a multimodal deep learning architecture for automated stress vulnerability assessment using Person in the Rain test drawings. The proposed method combines visual features extracted from drawings with textual descriptions for stress profiling as resilient or vulnerable. By integrating drawing characteristics and semantic narrative information, the study aims to provide a scalable alternative to traditional interpretation of projective drawing assessments. Results indicated that multimodal models highly outperformed image only approaches. ResNet50 + DistilBERT model achieving highest performance of 80% accuracy. Future work will include dataset expansion to increase generalizability, and exploration of advanced fusion strategies.

## ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176)

## REFERENCES

[1] E. Hammer, The Clinical Application of Projective Drawings. Springfield, IL, USA: Charles C Thomas, 1997.
[2] R. J. C. Koppitz, Psychological Evaluation of Human Figure Drawings. New York, NY, USA: Grune & Stratton, 1968.
[3] J. A. Lilienfeld, S. J. Wood, and J. M. Garb, “The scientific status of projective techniques,” Psychol. Sci. Public Interest, vol. 1, no. 2, pp. 27–66, 2000.
[4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. CVPR, 2016, pp. 770–778.
[5] A. Dosovitskiy et al., “An image is worth 16×16 words: Transformers for image recognition at scale,” in Proc. ICLR, 2021.
[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACL, 2019.
[7] A. Radford et al., “Learning transferable visual models from natural language supervision,” in Proc. ICML, 2021.
[8] J. Kim et al., “Developing an AI-based explainable expert support system for art therapy,” ACM Trans. Interact. Intell. Syst., vol. 14, no. 4, Art. 30, 2024.
[9] M. Yang et al., “CheckDAPR: An MLLM-based sketch analysis system for Draw-A-Person-in-the-Rain assessments,” in Proc. CIKM, 2025.
[10] A. I. Posada Barrera and M. O. Peña Diaz, “Building automated systems to generate structured data from Draw-A-Person-in-the-Rain test using deep learning,” in Proc. IEEE LA-CCI, 2025.
[11] Y. Chen et al., “Automatic dementia screening and scoring by applying deep learning on clock-drawing tests,” 2022.
[12] R. Wahyuningrum et al., “A computer vision system for automated scoring of hand-drawn geometric figures,” 2024.
[13] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, “Multimodal machine learning: A survey and taxonomy,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 2, pp. 423–443, 2019.
[14] J. Lu et al., “ViLBERT: Pretraining task-agnostic visiolinguistic representations,” in Proc. NeurIPS, 2019.
[15] H. Huang et al., “Fusion of medical imaging and clinical data for diagnosis: A deep learning approach,” IEEE J. Biomed. Health Inform., 2020.

# Deep Learning and Machine Learning based Framework for Sleep Stage Classification Using PPG Signals

*Awais Khan[1], Chomyong Kim[2], Yunyoung Nam[3, *]*

*[1]Department of ICT Convergence, Soonchunhyang University, Asan, 31538, Republic of Korea*

*[2] ICT Convergence Research Center, Soonchunhyang University, Asan, 31538, Republic of Korea*

*[3] Department of Computer Science and Engineering, Soonchunhyang University, Asan, South Korea*

**Contact: ynam@sch.ac.kr*

***Abstract*— Sleep is a fundamental physiological process essential for overall health and well-being, and its assessment traditionally relies on manual polysomnography (PSG), which is time-consuming, labor-intensive, and costly. Recent advances in Artificial Intelligence (AI) and wearable sensing technologies have enabled automated sleep stage classification using photoplethysmography (PPG) signals as a practical alternative for long-term monitoring. In this study, we utilized the Aalborg University Wearable Sleep Study (AAUWSS) dataset to develop a machine learning-based framework for classifying sleep stages, including Wake, N1, N2, N3, and REM, using PPG signals recorded from wearable devices. The proposed approach involves preprocessing and normalization of 30-second PPG epochs followed by classification using multiple machine learning models, including K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) variants. Experimental evaluation demonstrates that the Quadratic SVM achieved the best performance, attaining an accuracy of 87.5% and an ROC value of 0.87, outperforming other classifiers. These findings confirm the effectiveness of PPG-based sleep staging using machine learning techniques. Future work will focus on subject-independent validation and real-time wearable deployment.**

## I. INTRODUCTION

Sleep is a fundamental biological process that occupies nearly one-third of human life and is essential for maintaining physical health, cognitive function, and emotional stability [1]. In contemporary life, sleep disruptions have grown more prevalent despite their significance. A significant increase in insomnia symptoms was noted in large-scale population surveys carried out during the COVID-19 pandemic, underscoring the rising incidence of sleep-related issues globally [2]. Neurological and mental problems are often associated with sleep difficulties. For instance, in Alzheimer's patients, a decrease in deep sleep has been linked to a worsening of cognitive impairment [3]. In a similar vein, it has been demonstrated that persistent sleeplessness considerably raises the risk of depression, even in people who have never experienced depressive symptoms [4]. Therefore, early detection and treatment of sleep abnormalities may be a crucial preventive measure for more general mental health conditions. Accurate sleep architecture assessment is essential for prompt diagnosis and intervention given these clinical implications [5, 6].

Polysomnography (PSG), which records several physiological signals during nocturnal monitoring, such as electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG), is the clinical standard for assessing sleep [7]. In a procedure known as sleep stage classification or sleep scoring, these signals are divided into 30-second intervals, or epochs, and each epoch is allocated to a certain sleep stage. The Rechtschaffen and Kales (R&K) recommendations, which classified sleep into wakefulness (W), rapid eye movement (REM), and non-rapid eye movement (NREM), with NREM further subdivided into four substages (S1–S4), were the original definition of sleep staging criteria [8]. The American Academy of Sleep Medicine's (AASM) most recent standards combined stages S3 and S4 into a single stage, creating five widely used categories: W, N1, N2, and N3 as presented in Table 1. In clinical practice, scorers consider temporal continuity when assigning stages, as isolated epochs may be reclassified based on surrounding segments [9, 10]. These complexities highlight the challenges associated with reliable and automated sleep stage classification.

**Table 1. Sleep stages details.**

| Stages of sleep | No of classes | Abbreviations |
|---|---|---|
| REM, N1, N2, N3 and W | 6 | L means light sleep, that is N1 and N2 |
| REM, S1, S2, S3, S4 and W | 5 | D means deep sleep, that is N3 |
| REM, W, L and D | 4 | |
| REM, NREM, W | 3 | NREM and REM are considered as stages |
| W, Sleep | 2 | |

### A. Problem statement

- *The* Polysomnography (PSG) provides accurate sleep staging but is complex, expensive, and unsuitable for continuous home monitoring.
- Wearable alternatives such as PPG are more practical, but sleep stage classification using single-channel PPG remains challenging due to limited physiological information.

- Most existing studies rely on handcrafted HRV features and traditional machine learning, which may not fully capture nonlinear and temporal patterns in sleep signals.
- Fine-grained five-class sleep staging (Wake, N1, N2, N3, REM) using raw PPG signals is still insufficiently explored.

### *B. Major Contributions*

The objective of this study is to address the limitations of current methods by introducing a novel deep learning method for precise sleep stage classification. The proposed framework includes the following steps:

- Developing an automated sleep stage classification framework using single-channel PPG signals collected by Empatica E4, where 30-second epochs are preprocessed through z-score normalization to reduce inter-subject variability and improve signal consistency for reliable model training.
- Designing and implementing deep learning models, including CNN and LSTM architectures, to directly learn discriminative temporal features from raw PPG signals for fine-grained five-class sleep staging (Wake, N1, N2, N3, REM).
- Extracting deep representations from the learned feature layer and further evaluating their effectiveness using traditional machine learning classifiers such as SVM and KNN to enhance classification performance and provide comparative analysis.

## II. PROPOSED METHOD

In this work, we use single-channel PPG signals to present an automated approach for multi-class sleep stage classification, as shown in Figure 1. The proposed strategy is divided into four primary phases. PPG signals are gathered from several subjects in the first step and divided into 30-second intervals. In order to decrease inter-subject variability and enhance signal consistency, z-score normalisation and the elimination of epochs with missing labels are part of the second step of data pre-processing. Wake, N1, N2, N3, and REM are the five typical sleep stages that are used to name each epoch in the third stage. Using deep learning architectures, deep feature representations are directly learned from the normalised raw PPG signals in the fourth stage. The obtained features are given to machine learning classifiers for final classification after feature extraction. In particular, the discriminative power of the retrieved deep features is assessed using Support Vector Machine (SVM) and K-Nearest Neighbour (KNN) classifiers. Combining classical machine learning classifiers with deep learning-based representation learning improves classification robustness and offers a thorough assessment of sleep stage prediction ability. Figure 1 depicts the general design of the suggested framework.

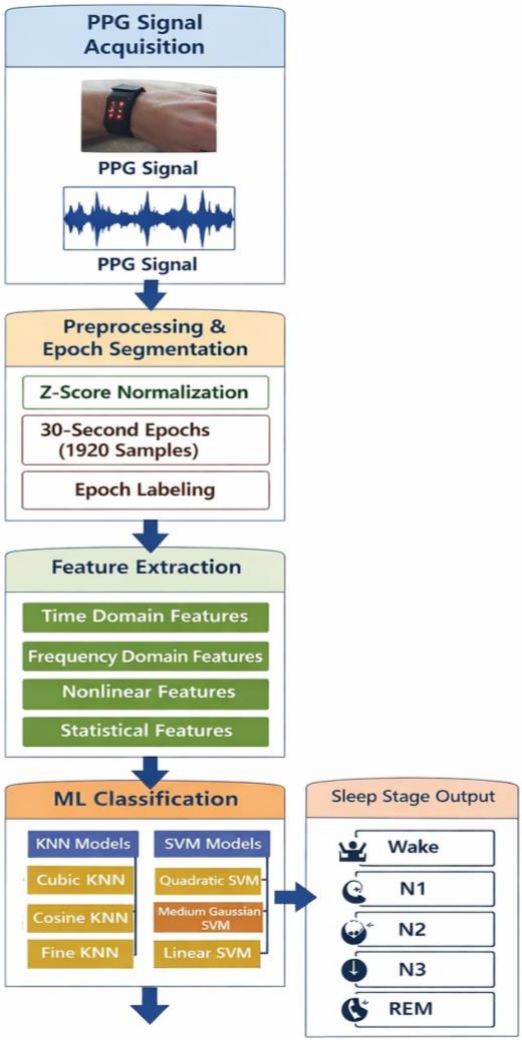



**Figure 1.** Proposed model architecture of sleep satge classification.

### *A. Data Description*

A multimodal sleep dataset titled the Aalborg University Wearable Sleep Study (AAUWSS) was collected from thirteen healthy adult participants during overnight sleep sessions at the Respiration Center West, Aarhus University Hospital [11]. Each participant wore a wearable Empatica E4 device in addition to a complete polysomnography (PSG) system, allowing for the simultaneous collection of wearable and clinical data. EEG, EOG, EMG, and ECG data from PSG recordings were manually assessed in accordance with the AASM sleep staging criteria to produce ground truth labels for each 30-second epoch, including Wake, N1, N2, N3, and REM stages.Although the synchronised and aligned PPG and ECG signals were the main emphasis for sleep stage classification, the Empatica E4 also recorded other physiological modalities, including skin temperature, accelerometer (ACC), electrodermal activity (EDA), and photoplethysmogram (PPG). Some subjects (e.g., Subjects 12 and 13) have patched aligned data because of PSG pauses during some recordings. The "aligned_sleep_data_set" aligns PPG and ECG signals with sleep stage labels; each row represents a labelled 30-second epoch collected at 64 Hz for PPG (1920 samples per epoch) and 200 Hz for ECG (6000 samples per epoch). The dataset also includes a participant information file containing

anonymized demographics such as age and gender, as well as auxiliary sensor recordings, providing a rich resource for wearable-based sleep stage research.

### *B. Signal processing and visualization*

The PPG data were obtained from the Aalborg University Wearable Sleep Study (AAUWSS) dataset and loaded from serialized .pkl files using Python. Each recording consists of 30-second epochs sampled at 64 Hz using the Empatica E4 device, resulting in 1920 samples per epoch. The first 1920 columns correspond to the raw PPG waveform, while the remaining columns contain sleep stage annotations. After loading the dataset, the PPG signals and their corresponding sleep stage labels were separated for further analysis, and the unique sleep stages present in the recording were identified. Data plotting with respect to window and wake label is shown in Figure 2.

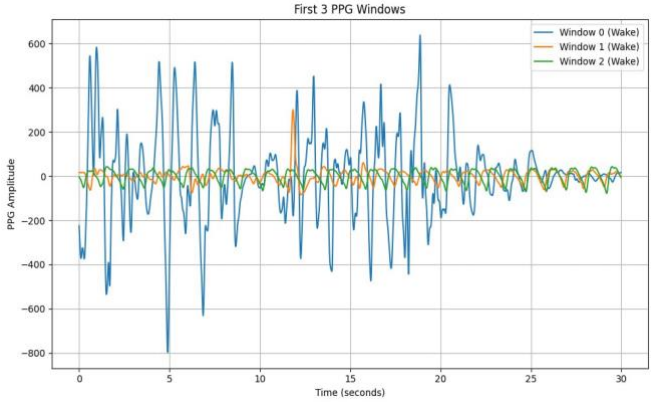



**Figure 2.** Data plotting with respect to window and wake label.

To visually analyse stage-specific characteristics, one representative 30-second epoch from each sleep stage was selected and plotted, as shown in Fig. X. For each stage (Wake, N1, N2, N3, and REM), the first occurrence in the dataset was extracted and displayed as a time-domain waveform in seconds. The subplots enable direct comparison of amplitude variations and morphological differences across sleep stages. These visualizations provide qualitative evidence of physiological variability in PPG signals and support the motivation for applying deep learning models for automatic sleep stage classification. Data plotting with respect to window and each class is shown in Figure 2.

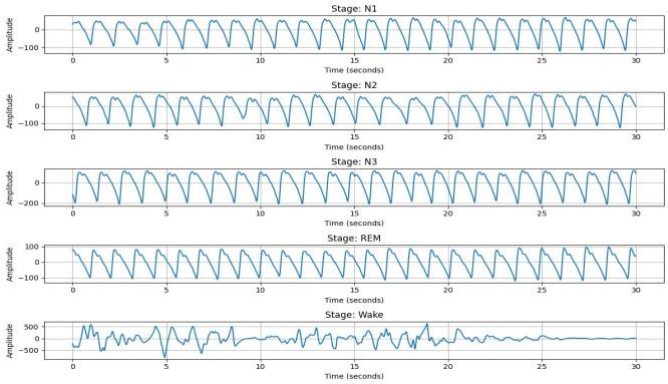



**Figure 3.** Data plotting with respect to window and each class.

### *C. Feature extraction using CNN*

The Convolutional Neural Network (CNN) model is designed to automatically learn discriminative temporal features from 30-second PPG epochs containing 1920 samples. Each epoch is first normalized using z-score normalization and reshaped into a 1920 × 1 × 1 format to match the network input structure. The architecture consists of three convolutional blocks with 32, 64, and 128 filters, respectively, using kernel sizes of 5, 5, and 3 with same padding. Each convolutional layer is followed by batch normalization and ReLU activation to stabilize training and introduce non-linearity, while max-pooling layers are applied after the first two convolutional blocks to reduce temporal dimensionality. The extracted feature maps are passed to a fully connected layer with 128 neurons, followed by ReLU activation and a dropout layer with a rate of 0.5 to reduce overfitting. The final classification layer uses a softmax activation function to predict the five sleep stages (Wake, N1, N2, N3, REM). The network is trained using the Adam optimizer with cross entropy loss.

### *D. Feature extraction using LSTM*

In The LSTM model is employed to capture long-range temporal dependencies within each 30-second PPG epoch, treated as a sequence of length 1920 with one feature per time step. After z-score normalization, each epoch is fed into a sequence input layer followed by a single LSTM layer with 128 hidden units configured in “last” output mode, allowing the final hidden state to represent the entire sequence. This representation is passed to a fully connected layer with 128 neurons, followed by a ReLU activation function and a dropout layer with a rate of 0.5 to prevent overfitting. A final fully connected layer and softmax function perform classification into the five sleep stages. The model is trained using the Adam optimizer and optimized with cross-entropy loss, enabling effective modeling of sequential physiological patterns in the PPG signal.

## III. Results and Discussion

This section presents the experimental results of the proposed CNN and LSTM along with machine learning based method. The dataset was split into 80:20, 80% data for training and 20% for testing. The training parameters includes, 50 iterations, 40 epochs, a minibatch size of 32, and Adam optimizer. However, the extracted features extracted from the CNN and LSTM were used as an input to six machine learning classifiers, including linear SVM, Quadratic SVM, Medium Gaussian SVM, Cubic KNN, Cosine KNN, Fine KNN. Each model was evaluated using training time, accuracy, recall, ROC, precision rate, and confusion matrices to assess classification performance. Experiments were conducted using Python on a high-performance system.

### *A. Results for CNN and ML models*

The performance of the evaluated machine learning models is summarized in Table 2. Among all classifiers, the Quadratic Support Vector Machine (QSVM) achieved the highest accuracy of 87.5%, along with a precision rate of 85.9%, recall of 88.3%, and an ROC value of 0.85, while maintaining a relatively low training time of 4.5 seconds. The Medium Gaussian SVM (MGSVM) also demonstrated strong performance with an accuracy of 83.9% and ROC of 0.83, although with higher computational cost. The Linear

SVM (LSVM) achieved moderate performance with 80.6% accuracy and an ROC value of 0.80. In comparison, KNN-based models showed comparatively lower classification performance. Cubic KNN and Cosine KNN achieved accuracies of 75.1% and 74.7%, respectively, while Fine KNN produced the lowest accuracy of 70.8%. Overall, SVM-based classifiers outperformed KNN variants in terms of accuracy, precision, recall, and ROC performance, indicating that margin-based learning methods are more effective for sleep stage classification using the extracted features.

**Table 2.** Experimental results for sleep stage classification CNN and machine learning.

| ML Models | Accuracy (%) | Training Time (s) | Precision Rate | Recall Rate | ROC |
|---|---|---|---|---|---|
| Cubic KNN | 75.1 | 6.1 | 74.5 | 78.5 | 0.80 |
| Cosine KNN | 74.7 | 8.2 | 69.6 | 77.5 | 0.78 |
| Fine KNN | 70.8 | 6.3 | 69.8 | 70.2 | 0.75 |
| QSVM | 87.5 | 4.5 | 85.9 | 88.3 | 0.85 |
| MGSVM | 83.9 | 10.1 | 81.3 | 83.5 | 0.83 |
| LSVM | 80.6 | 8.3 | 76.8 | 79.8 | 0.80 |

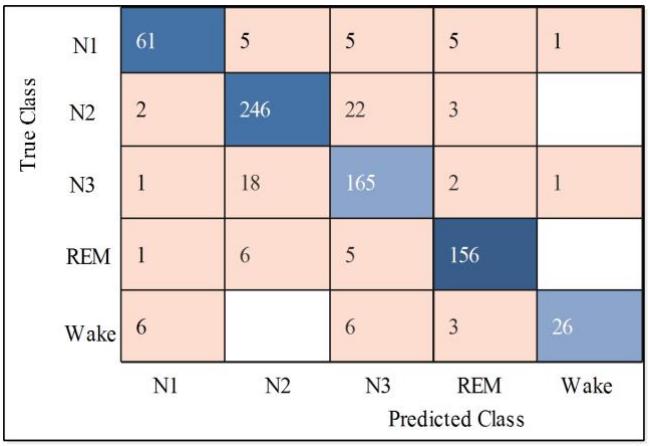



**Figure 4.** Confusion matrix for sleep stage classification using CNN and QSVM.

*B. Experimental results using LSTM and machine learning*

The performance comparison of the evaluated machine learning classifiers is presented in Table X. Among all models, the Medium Gaussian SVM (MGSVM) achieved the highest accuracy of 86.9% with an ROC value of 0.83, demonstrating strong generalization capability despite a moderate training time of 9.1 seconds. The Quadratic SVM (QSVM) also showed competitive performance with an accuracy of 82.5%, the highest precision (84.1%) and recall (84.5%), and a relatively low training time of 3.8 seconds, making it computationally efficient. The Linear SVM (LSVM) achieved 78.9% accuracy with an ROC of 0.79. In contrast, KNN-based models exhibited comparatively lower performance. Cubic KNN and Cosine KNN achieved accuracies of 73.5% and 73.1%, respectively, while Fine KNN produced the lowest accuracy of 69.2% and an ROC of 0.70. Overall, SVM-based classifiers outperformed KNN variants in terms of accuracy, precision, recall, and ROC performance, indicating that kernel-based decision boundaries are more effective for sleep stage classification using the extracted feature representations.

**Table 3.** Experimental results for sleep stage classification using LSTM and machine learning.

| ML Models | Accuracy (%) | Training Time (s) | Precision Rate | Recall Rate | ROC |
|---|---|---|---|---|---|
| Cubic KNN | 73.5 | 7.4 | 72.4 | 77.7 | 0.78 |
| Cosine KNN | 73.1 | 8.5 | 68.2 | 75.3 | 0.77 |
| Fine KNN | 69.2 | 7.6 | 67.9 | 68.4 | 0.70 |
| QSVM | 82.5 | 3.8 | 84.1 | 84.5 | 0.80 |
| MGSVM | 86.9 | 9.1 | 80.4 | 81.6 | 0.83 |
| LSVM | 78.9 | 12.5 | 76.2 | 79.3 | 0.79 |

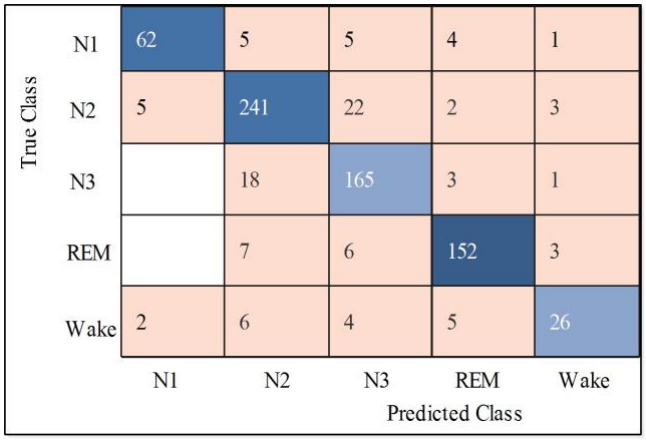



**Figure 5.** Confusion matrix for sleep stage classification using CNN and QSVM.

## IV. Conclusion

This This study presented a machine learning-based framework for automatic sleep stage classification using single-channel PPG signals. The extracted features were evaluated using multiple classifiers, including KNN and SVM variants. Experimental results demonstrated that SVM-based models significantly outperformed KNN classifiers in terms of accuracy, precision, recall, and ROC performance. In particular, the Medium Gaussian SVM achieved the highest classification accuracy of 86.9% with strong discriminative capability (ROC = 0.83), while the Quadratic SVM provided competitive performance with lower computational cost. These findings indicate that kernel-based classifiers are well-suited for modeling the nonlinear characteristics of PPG-derived sleep features. Overall, the results confirm the feasibility of using wearable PPG signals for reliable sleep stage classification.

Although promising results were achieved, several improvements can be explored in future research. First, advanced deep learning architectures such as hybrid CNN–LSTM or transformer-based models may further enhance temporal feature learning. Second, subject-independent validation strategies such as leave-one-subject-out cross-validation should be investigated to improve generalizability. Additionally, addressing class imbalance through weighted loss functions or data augmentation

techniques could improve minority stage detection, particularly N1 sleep. Finally, integrating multimodal physiological signals, such as accelerometer or electrodermal activity data from wearable devices, may further enhance classification robustness and real-world applicability. In future work, we plan to perform the proposed method on our self-collected dataset and conduct a comprehensive comparison with state-of-the-art (SOTA) methods to further validate its efficiency and robustness.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176).

## REFERENCE

[1] Aboalayon KAI, Faezipour M, Almuhammadi WS et al (2016) Sleep stage classifcation using EEG signal analysis: a comprehensive survey and new investigation. Entropy 18(9):272. https://doi.org/10.3390/e18090272

[2] Adib F (2019) Seeing with radio wi-f-like equipment can see people through walls, measure their heart rates, and gauge emotions. IEEE Spectr 56(6):34–39. https://doi.org/10.1109/MSPEC.2019.8727144

[3] Afonso VX, Tompkins WJ, Nguyen TQ et al (1999) Ecg beat detection using flter banks. IEEE Trans Biomed Eng 46(2):192–202. https://doi.org/10.1109/10.740882

[4] Ali PJM, Faraj RH, Koya E et al (2014) Data normalization and standardization: a technical report. Mach Learn Tech Rep 1(1):1–6

[5] Al-Saegh A, Dawwd SA, Abdul-Jabbar JM (2021) Deep learning for motor imagery EEG-based classifcation: a review. Biomed Signal Process Control 63:102172. https://doi.org/10.1016/j.bspc.2020.102172

[6] Alsolai H, Qureshi S, Iqbal SMZ et al (2022) A systematic review of literature on automated sleep scoring. IEEE Access 10:79419–79443

[7] Altaheri H, Muhammad G, Alsulaiman M et al (2023) Deep learning techniques for classifcation of electroencephalogram (EEG) motor imagery (MI) signals: a review. Neural Comput Appl 35(20):14681–14722. https://doi.org/10.1007/s00521-021-06352-5

[8] Arif S, Khan MJ, Naseer N et al (2021) Vector phase analysis approach for sleep stage classifcation: a functional near-infrared spectroscopy-based passive brain-computer interface. Front Hum Neurosci 15:658444

[9] Baek J, Lee C, Yu H et al (2022) Automatic sleep scoring using intrinsic mode based on interpretable deep neural networks. IEEE Access 10:36895–36906. https://doi.org/10.1109/ACCESS.2022.3163250

[10] Baglioni C, Battagliese G, Feige B et al (2011) Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Afect Disord 135(1–3):10–19. https://doi.org/10. 1016/j.jad.2011.01.011.

[11] Djanian, S., Nielsen, T. D., Nielsen, S. H., & Bruun, A. (2025). Aalborg University Wearable Sleep Study (AAUWSS) (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.16919071

# *Development and Feasibility of a VR-Based Cognitive Assessment Tool for Early Screening of Developmental Delay in Preschool Children*

Chomyong Kim[1], Yeana Ha[1], Joungmin Kim[2], Yunyoung Nam[1,3], and Eun Young Kim[4]

[1]*Emotional and Intelligence Child Care System Convergence Research Center, Soonchunhyang University, Asan, Republic of Korea*

[2]*Dept. of Liberal Arts, , Soonchunhyang University, Asan, Republic of Korea*

[3]*Dept. of Computer Science and Engineering, 4Dept. of Occupational Therapy, Soonchunhyang University, Asan, Republic of Korea*

**Contact: eykim@sch.ac.kr; ynam@sch.ac.kr*

***Abstract*— Early identification of developmental delay (DD) in preschool children is critical for optimizing developmental outcomes, yet standardized immersive VR-based cognitive assessment tools for children aged 3–5 years have not been systematically developed. This pilot study developed six VR cognitive tasks assessing Object Concept, Color Concept, and Numerical Understanding using the Meta Quest Pro HMD with controller-free hand tracking, and implemented an AABB-based automatic scoring system. Fifty-four children (DD: n=17, ND: n=37; mean age 4.63±0.91 years) participated. A mean task execution rate of 83.6% confirmed technical feasibility. Four tasks demonstrated excellent discriminability (D≥0.40) and appropriate difficulty (accuracy 40–65%). Age-adjusted ANCOVA revealed no significant group difference in VR total scores ($F(1,51)=0.30$, $p=.584$), whereas age was a significant predictor ($F(1,51)=9.50$, $p=.003$, $\eta^2=0.156$). These findings support the feasibility of immersive HMD-based VR cognitive assessment for preschool children and indicate psychometric potential of four tasks as early screening indicators.**

## I. Introduction

Virtual reality (VR) technology has demonstrated significant potential across healthcare, education, and cognitive assessment by providing immersive, standardized, and ecologically valid testing environments [1]. In pediatric healthcare, VR-based tools have been applied to assess and intervene in attention deficit hyperactivity disorder [2], autism spectrum disorder [3], and a range of neurodevelopmental conditions [4][5][6].

Despite growing evidence, most pediatric VR assessment studies have targeted children aged five years and older [7]. Early childhood (ages 0–5) is a critical window for cognitive development; undetected delays can escalate into more severe intellectual disabilities [8]. Tablet-based tools have demonstrated feasibility for early cognitive screening [9], but immersive HMD-based VR offers distinct advantages through three-dimensional, controller-free interaction that more closely approximates real-life assessment conditions [10].

To the best of our knowledge, no prior study has systematically developed and evaluated immersive HMD-based VR cognitive assessment specifically targeting children aged 3 to under 6 years. This study aimed to: (1) develop VR cognitive assessment content using the Meta Quest Pro HMD with controller-free hand tracking; (2) implement an automatic scoring system for objective, examiner-independent evaluation; and (3) evaluate the technical feasibility and initial psychometric properties of the system in a pilot sample.

## II. System Deployment

### *A. Hardware and Software Environment*

VR content was developed using Unity 2022.3.62f2 and deployed on the Meta Quest Pro HMD (resolution: 1,800×1,920 px/eye; FOV: 106°×96°; refresh: 90 Hz; eye-tracking: up to 50 Hz). Scripts were written in C# using Visual Studio 2022. 3D modeling was performed in Blender 3.6, NPC detailing in ZBrush 2023, and rigging and facial animation in Autodesk Maya 2024. Motion capture was recorded at 120 Hz using Perception Neuron sensors and Axis Studio. Content was integrated via the OVR Plugin (Meta Quest Pro SDK, April 2024).

### *B. Cognitive Task Design*

Six VR cognitive tasks were developed across three constructs based on established developmental milestones for children aged 36–71 months [11] and a prior expert committee process [12]: (1) Object Concept — Task 1 (Object Memory), Task 2 (Similarity Identification), Task 3 (Object Classification); (2) Color Concept — Task 4 (Color Sorting), Task 6 (Color Classification); (3) Numerical Understanding — Task 5 (Counting). Internal development labels (Tasks 1, 3, 4, 5, 6, 7) differ from paper labels (Tasks 1–6) as the internal Task 2 (picture naming) was excluded due to ceiling effects.

An NPC-based virtual assessor modeled after a typical Korean daycare teacher was developed to deliver standardized instructions, replacing a human examiner. Motion capture was performed by a licensed occupational therapist (Ph.D.) to ensure procedural accuracy. A total of 32 animation clips (~15 s each, 3–5 per task) were produced and synchronized with pre-recorded Korean voice audio. The virtual environment simulated a daycare setting with controlled visual stimulation (light beige/gray tones, blocked windows, overhead task lighting).

### C. Controller-Free Hand Tracking Interaction

A hand-tracking interaction module was implemented via the April 2024 Meta Quest Pro SDK, enabling direct finger-based object manipulation without controllers. The Distance Grab feature was disabled to restrict interaction to direct contact only. Collider sizes exceeded physical object dimensions to ensure stable contact recognition given children's hand size. Interaction zones were configured to auto-return objects displaced beyond a set range, preventing assessment interruption.

### D. Automatic Answer Evaluation

Task-specific scoring rules were predefined in an id_mapping.json file and applied at runtime using AABB (Axis-Aligned Bounding Box) collision detection in Unity 3D space. A collision sustained for ≥0.5 seconds constituted a correct response. Three rule types were applied: exclusive_group (Tasks 1, 3, 5, 6), standard (Task 2), and standard + empty_zone (Task 4). Scores and timestamps were automatically logged to CSV without examiner involvement, eliminating inter-rater variability.

### E. Data Preparation Pipeline

A two-stage Python pipeline (pandas, numpy, matplotlib) was developed for raw data processing. Stage 1 (session segmentation) used the scene_act variable to extract task-level data intervals; sessions with execution time <10 seconds were excluded at the participant level. Stage 2 (3D spatial validation) applied an interactive visualization tool with MAD-based outlier detection (MAD_K=6.0) to identify non-activated objects and confirm accurate task boundary intervals.

## III. Study Design

### A. Participants

This prospective pilot study enrolled children aged 36–71 months recruited from daycare centers, kindergartens, hospitals, and therapy centers (June 2024–January 2026, Soonchunhyang University Research Center). Inclusion criteria: (1) age 36–71 months, (2) written guardian consent, (3) ability to wear the HMD. Exclusion criteria: visual/auditory impairments precluding VR recognition, severe motor disabilities preventing hand-tracking interaction, history of epilepsy or photosensitive seizures.

Of 110 recruited children, 29 were unable to participate (HMD refusal/adaptation failure). Of 81 who participated, 9 were excluded for age non-compliance, 1 for execution time <10 seconds, and 16 for early session termination, yielding N=54 (DD: n=17, ND: n=37). Group classification was performed by a licensed occupational therapist using the Korean Child Development Inventory (K-CDI): DD defined as total T-score ≤30 (≥2 SD below mean), ND as T-score >30.

### B. Measures and Statistical Analysis

Technical feasibility was evaluated via task execution rate and accuracy rate. Psychometric properties were assessed using difficulty index and discriminability index (D; Ebel & Frisbie, 1991 [14]): Excellent (D≥0.40), Good (0.30≤D<0.40), Acceptable (0.20≤D<0.30), Poor (D<0.20). Internal consistency was assessed using KR-20. Between-group differences in total VR scores were analyzed via ANCOVA with age as covariate; task-level comparisons used Fisher's exact test or chi-square with phi coefficient (φ). Exploratory ROC analysis was performed using the VR total score as a screening indicator.

### C. Ethics

The study was approved by the Institutional Review Board of Soonchunhyang University (No. 1040875-202401-SB-005; No. 1040875-2025-03-022, reflecting annual renewal of a single approval). Written informed consent was obtained from guardians; age-appropriate verbal assent was obtained from children.

## IV. Results

### A. Participant Characteristics

Demographic characteristics are summarized in Table II. The DD group was significantly younger than the ND group (4.27±0.83 vs. 4.80±0.90 years; t(52)=−2.035, p=.047, d=0.60). Sex distribution did not differ significantly (Fisher's exact p=.364). The mean developmental age of the DD group (41.9±14.7 months) was below chronological age, consistent with mild-to-moderate delay.

TABLE I.
DEMOGRAPHIC CHARACTERISTICS OF PARTICIPANTS (N=54)

| Variable | DD (n=17) | ND (n=37) | Statistic |
|---|---|---|---|
| Age, M±SD (yrs) | 4.27±0.83 | 4.80±0.90 | t(52)=−2.04, p=.047 |
| Male, n (%) | 13 (76.5) | 23 (62.2) | Fisher p=.364 |
| Dev. Age (mo), M±SD | 41.9±14.7 | 58.1±10.2 | — |
| VR Total Score, M±SD | 2.29±1.69 | 2.92±1.55 | t p=.187 |

### B. Technical Feasibility and Psychometric Properties

Task performance outcomes are summarized in Table I. The mean execution rate of 83.6% confirmed technical feasibility of controller-free hand-tracking interaction. KR-20 was 0.68, considered acceptable for a multidimensional pilot battery. Task 1 exhibited a ceiling effect (accuracy 94.2%, D=0.17; Poor). Task 6 exhibited a floor effect (accuracy 12.2%, D=0.27; Acceptable). Tasks 2 (59.6%, D=0.87), 3 (40.9%, D=0.87), 4 (44.2%, D=0.67), and 5 (63.6%, D=0.81) demonstrated appropriate difficulty and Excellent discriminability. Task 2 was the only task with a significant between-group difference in non-attempt rate (DD: 29.4% vs. ND: 5.4%; Fisher's p=.026), suggesting task avoidance as an additional behavioral indicator.

TABLE II.

TASK PERFORMANCE OUTCOMES AND PSYCHOMETRIC PROPERTIES (N=54)

| Task | Exec. | Acc. | Diff. | D | Interp. |
|---|---|---|---|---|---|
| 1 (Obj.Memory) | 96.3% | 94.2% | 0.06 | 0.17 | Poor |
| 2 (Similarity) | 87.0% | 59.6% | 0.40 | 0.87 | Excellent |
| 3(Classification) | 81.5% | 40.9% | 0.59 | 0.87 | Excellent |
| 4 (Color Sort) | 79.6% | 44.2% | 0.56 | 0.67 | Excellent |
| 5 (Counting) | 81.5% | 63.6% | 0.36 | 0.81 | Excellent |
| 6 (Color Class.) | 75.9% | 12.2% | 0.88 | 0.27 | Acceptable |

D = discriminability index (upper 27% − lower 27%; Ebel & Frisbie, 1991). Exec. = Execution rate; Acc. = Accuracy rate; Diff. = Difficulty index; Interp. = Interpretation.

### C. Group Comparison

ANCOVA assumptions were satisfied (linearity: r=0.425, p=.001; homogeneity of regression slopes: F=1.51, p=.226). After controlling for age, no significant between-group difference was found in VR total scores ($F(1,51)=0.30$, $p=.584$, $\eta^2=0.005$). Age was a significant main effect ($F(1,51)=9.50$, $p=.003$, $\eta^2=0.156$). Adjusted means at covariate mean age 4.63 years were 2.55 (DD) and 2.80 (ND) points. The supplementary t-test was also non-significant ($t(52)=-1.337$, $p=.187$, $d=0.392$; 95% CI [−0.31, 1.56]). Post-hoc power analysis indicated observed power of 25.9% for $d=0.392$ at $\alpha=.05$ with N=54, well below the recommended 80%; N=208 would be required. At the task level, Task 2 showed the largest gap (+24.0%p, $\varphi=.21$) and Task 4 showed +15.1%p ($\varphi=.14$), both non-significant.

### D. Exploratory Screening Analysis

ROC analysis using the VR total score (range 0–6) yielded AUC=0.61 (95% CI [0.431, 0.773]), exceeding chance level (0.50). The optimal cutoff by Youden's J was ≤1 (sensitivity 47.1%, specificity 75.7%, J=0.227), providing preliminary evidence of screening potential pending larger-sample validation.

## V. DISCUSSION

This pilot study is, to our knowledge, the first to systematically develop and report psychometric properties of an immersive HMD-based VR cognitive assessment for preschool children aged 3 to under 6 years. A mean execution rate of 83.6% demonstrates that controller-free hand-tracking interaction is technically viable for this age group. The non-significant ANCOVA result should be interpreted in the context of severely limited statistical power (25.9%) rather than as evidence of no group effect. The observed effect size ($d=0.392$) and ROC AUC of 0.61 indicate clinically relevant variance, and performance is expected to improve substantially with age-stratified normative data and larger samples. Age explaining 15.6% of VR score variance ($\eta^2=0.156$) further underscores the necessity of age-stratified scoring in a future validated instrument.

Compared to prior mobile touch-based approaches for preschool cognitive screening [9], the present system advances the field by transitioning to an immersive HMD format, implementing controller-free hand tracking, and enabling fully automatic examiner-independent scoring via 3D AABB collision detection. Tasks 1 and 6 require redesign to address ceiling and floor effects, respectively. The four remaining tasks provide a psychometrically sound foundation for a validated battery. Limitations include the small single-site sample, K-CDI-only group classification, and the absence of multimodal behavioral data analysis (hand-tracking kinematics, gaze, head motion), which represent major avenues for future machine learning-based screening model development.

## VI. CONCLUSION

This study demonstrates the technical feasibility of immersive HMD-based VR cognitive assessment with controller-free hand tracking for preschool children aged 3 to under 6 years. Four of six developed tasks showed appropriate difficulty and excellent discriminability, supporting their potential as developmental delay screening indicators. Future work should prioritize multi-site validation with larger age-stratified samples, redesign of Tasks 1 and 6, and integration of multimodal behavioral data for machine learning-based early screening.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00218176) and the Soonchunhyang University Research Fund.

## REFERENCES

[1] P. M. G. Emmelkamp and K. Meyerbröker, "Virtual reality therapy in mental health," Annu. Rev. Clin. Psychol., vol. 17, no. 1, pp. 495–519, 2021.

[2] A. Bashiri, M. Ghazisaeedi, and L. Shahmoradi, "The opportunities of virtual reality in the rehabilitation of children with ADHD," Korean J. Pediatr., vol. 60, no. 11, pp. 337–343, 2017.

[3] R. Bradley and N. Newbutt, "Autism and virtual reality head-mounted displays: A systematic review," J. Enabling Technol., vol. 12, no. 3, pp. 101–113, 2018.

[4] P. Kourtesis et al., "Validation of the Virtual Reality Everyday Assessment Lab (VR-EAL)," J. Int. Neuropsychol. Soc., vol. 27, no. 2, pp. 181–196, 2021.

[5] M. L. Lorusso et al., "Semi-immersive virtual reality as a tool to improve cognitive and social abilities in preschool children," Appl. Sci., vol. 10, no. 8, p. 2948, 2020.

[6] T. Segawa et al., "VR in assessment and treatment of addictive disorders: A systematic review," Front. Neurosci., vol. 13, p. 1409, 2020.

[7] E. Seesjärvi et al., "EPELI: A novel VR task for the assessment of goal-directed behavior," Psychol. Res., vol. 87, no. 6, pp. 1899–1916, 2023.

[8] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Arlington, VA: American Psychiatric Publishing, 2013.

[9] H. J. Kim et al., "Detection of developmental disabilities in preschool-aged children through game-based assessment using machine learning," JMIR Serious Games, vol. 9, no. 2, p. e23130, 2021.

[10] T. D. Parsons, "Virtual reality for enhanced ecological validity and experimental control," Front. Hum. Neurosci., vol. 9, p. 660, 2015.

[11] Y. T. Kim and M. J. Shin, Korean Child Development Inventory (K-CDI). Seoul: Special Education, 2006.

[12] M. Hong, E. Y. Kim, Y. Nam, C. Kim, and J. Kim, "Development of metaverse assessment items for early screening of developmental disabilities," J. Korean Acad. Sensory Integr., vol. 22, no. 2, pp. 41–54, 2024.

[13] L. Crocker and J. Algina, Introduction to Classical and Modern Test Theory. New York: Holt, Rinehart and Winston, 1986.

[14] R. L. Ebel and D. A. Frisbie, Essentials of Educational Measurement, 5th ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.

# A Preliminary Study for Developing Metaverse-Based Language Intervention Contents for School-Aged Children with Language Disorders: Priorities for Intervention Situations

Youn Jin Ji[1], Soo-Hyang Lee [2], Ji Young Na [2], Gyung-Hun Hong [2]*
[1] *Department of Communication Disorders, The Graduate School of Korea Nazarene University, Cheonan, Korea*
[2]*Professor, Department of Communication Disorders, Korea Nazarene University, Cheonan, Korea*
*Corresponding Author: ghun@kornu.ac.kr

***Abstract*— The purpose of this study was to investigate the priority of intervention situations considered important by school-aged children with language disorders, their caregivers, and speech-language pathologists (SLPs). when implementing metaverse-based language intervention. A survey method was employed, and the survey items were developed based on a functional approach to language intervention, emphasizing the generalization of communication skills in everyday contexts. Six major life domains were identified: home, community and leisure, school environment, peer interaction, classroom activities, and career and vocational situations. A total of 177 participants, including 60 children, 57 caregivers, and 60 SLPs, participated in the study. The results showed that caregivers and SLPs demonstrated relatively similar response patterns, whereas children exhibited different preferences. The only situation ranked first by all three groups was using a convenience store in the community and leisure domain. Caregivers and SLPs tended to prioritize functional situations related to academic performance and daily adaptation, while children emphasized situations related to personal interests and real-life participation. These findings suggest that metaverse-based language intervention goals should be determined through shared decision making that incorporates the perspectives of children, caregivers, and professionals.**

## I. INTRODUCTION

This study aims to investigate the prioritization of intervention situations considered important by school-aged children with language disorders, their caregivers, and SLPs when implementing metaverse-based language intervention. [1] In particular, this study seeks to identify how different stakeholders perceive and prioritize functional communication situations in everyday contexts and to explore how these priorities differ across groups. Understanding these differences is essential for designing intervention goals that are both clinically effective and personally meaningful for children.

Recently**, metaverse environments based on virtual reality (VR) technology** have gained attention in SLP as an innovative approach that can overcome the limitations of time and space in clinical settings. VR-based environments enable users to engage in immersive and interactive experiences that closely resemble real-life situations, thereby providing opportunities for repeated practice and contextualized learning. Furthermore, the use of metaverse environments may facilitate the generalization of communication skills by allowing children to practice language in dynamic and meaningful contexts.

This is particularly important for children with language disorders, who often experience difficulties transferring learned skills to real-world situations. By embedding intervention within functional and engaging scenarios, metaverse-based approaches have the potential to enhance motivation, participation, and functional outcomes.

Despite these advantages, there is still limited empirical research on the prioritization of intervention situations when designing metaverse-based language intervention content.

Therefore, this study aims to address this gap by examining the prioritization of intervention situations across these groups and providing practical implications for the development of user-centered and functionally relevant metaverse-based language intervention programs.

## II. METHOD

This study employed a survey methodology to inform the development of metaverse-based language intervention content. The survey design was theoretically grounded in the functional approach to language intervention, which emphasizes the natural generalization of communicative competence within an individual's daily life by shifting the focus from structured clinical settings to authentic, real-world contexts.

To capture the multifaceted nature of a child's life, the researchers categorized the primary life domains of school-aged children into several key areas, including home, community and leisure, school environment, peer interactions, classroom activities, and career or vocational situations. Each domain was further operationalized into specific communicative scenarios reflecting various communicative challenges encountered by children in their everyday environments.

To ensure the academic rigor of the instrument, the survey underwent a content validity assessment by an expert panel to confirm the appropriateness of the items before final administration. The data collection process was conducted through both online and paper-based formats to accommodate the diverse characteristics of the participant groups. While SLPs and caregivers participated via an online platform accessible on mobile devices to maximize accessibility, child participants completed a paper-based version of the survey to facilitate better comprehension and engagement. During the administration for children, SLPs provided supplementary verbal explanations regarding the context of each scenario to ensure the reliability of the responses.

All respondents were tasked with reviewing the situations within each life domain and identifying the most critical scenarios for implementation in a metaverse-based intervention, thereby eliciting the core communicative contexts that stakeholders perceive as most essential for virtual reality-based therapeutic applications.

| | Domain | Definition |
|---|---|---|
|  | Home | Communication activities with family members at home |
|  | Community and Leisure | Communication activities in public spaces and facilities |
|  | School Environment | Communication activities for using various school facilities |
|  | Peer Interaction | Communication activities with peers |
|  | Classroom Activities | Communication activities during classroom instruction |
|  | Career/ Vocational Training Situations | Communication activities for career exploration and vocational skill-building |

TABLE I. Daily Life Situation Domains Included in the Questionnaire

## III. RESULTS

A total of 177 participants enrolled in the study, including 60 school-aged children with language disorders, 57 caregivers, and 60 SLPs. Overall, the response patterns of caregivers and SLPs were highly similar across domains, whereas children demonstrated relatively distinct preference patterns.

Across all domains, the only situation that was consistently ranked first by all three groups was *convenience store use* within *the community and leisure* domain (Children: 51.7%, Caregivers: 70.2%, SLPs: 68.3%). This finding indicates a shared recognition among stakeholders regarding the importance of functional, real-life communication contexts involving everyday transactions and social interactions in the community.

In the *home* domain, children prioritized *simple food preparation* (71.7%) and *pet care* (43.3%), reflecting their interest in concrete, hands-on activities. In contrast, caregivers and SLPs placed greater emphasis on functional daily living skills, such *as after-school routines* and *self-care*. A notable discrepancy emerged regarding *school preparation* and structured routines; while these were identified as essential by both adult groups, children showed a relatively lower prioritization of such organizational activities.

In the *community and leisure* domain, aside from the universally prioritized *convenience store use*, children ranked *kiosk use* and *family travel* as secondary priorities, indicating a preference for dynamic, experience-oriented activities. In contrast, caregivers and SLPs emphasized structured and safety-related contexts, such *as public transportation use*, *traffic rule compliance*, and *emergency reporting*, reflecting their concern for safety and socially appropriate behavior in public settings.

In the *school environment* domain, caregivers and SLPs showed highly similar response patterns, prioritizing *classroom rules compliance* (Caregivers: 100%, SLPs: 95.0%), *school cafeteria use*, and *restroom use* as key situations. Children, in contrast, ranked *school cafeteria use* (63.3%) and *playground use* (56.7%) as their top priorities, demonstrating a preference for activity-based and socially engaging contexts.

In the *peer interaction* domain, caregivers and SLPs prioritized *group activity participation* (Caregivers: 94.7%, SLPs: 93.3%) as the most important situation, reflecting an emphasis on cooperative communication and social integration. In contrast, children selected *birthday party participation* (63.3%) as their top priority, suggesting a stronger interest in socially meaningful and enjoyable events. Secondary priorities for caregivers and SLPs included *borrowing and returning items* and *meeting friends by appointment*, while children also showed interest in *board game participation*, highlighting engagement in structured play contexts.

In the *classroom activities* domain, caregivers and SLPs consistently emphasized academically oriented tasks such as *class preparation* (Caregivers: 84.2%, SLPs: 83.3%) and *delivering school announcements* (Caregivers: 64.9%, SLPs: 71.7%). In contrast, children prioritized *arts and physical education* (76.7%), indicating a preference for more active and less academically demanding classroom situations.

Finally, in the *career and vocational training* domain, caregivers and SLPs showed similar patterns, prioritizing structured, responsibility-based tasks such as *item sorting* and *job interviews*. Children, however, placed relatively higher importance on *coffee making* and *item assembly*, reflecting an interest in concrete, task-oriented activities rather than abstract or evaluative situations.

Taken together, these findings reveal a clear divergence in priority patterns between children and adult stakeholders. Caregivers and SLPs tended to prioritize situations directly related to academic performance, rule compliance, independence, and safety. In contrast, children consistently prioritized situations that were activity-based, socially engaging, and directly tied to their own interests and lived experiences. These differences highlight the importance of incorporating multiple perspectives, particularly those of children, when designing metaverse-based language intervention content to ensure both functional relevance and user engagement.

| Domain | Rank | Children N= 60 Frequency(%) | Caregivers N= 57 Frequency(%) | SLPs N= 60 Frequency(%) | Total (%) |
|---|---|---|---|---|---|
| Home | 1 | Simple Food Preparation 43(71.7) | Daily Routines after School 45(78.9) | self-care skills 52(86.7) | 79.1% |
| | 2 | Pet Care 26(43.3) | self-care skills 44(77.2) | School Preparation 43(71.7) | 64.1% |
| | 3 | Device Use at Home 17(28.3) | School Preparation 34(59.6) | Daily Routines after School 38(63.3) | 50.4% |
| Community and Leisure | 1 | Convenience Store Use 31(51.7) | Convenience Store Use 40(70.2) | Convenience Store Use 41(68.3) | 63.4% |
| | 2 | Kiosk Use 27(45.0) | Public Transportation Use 32(56.1) | Traffic Rule Compliance 39(65.0) | 55.4% |
| | 3 | Family Travel 23(38.3) | Emergency Reporting 32(56.1) | Public Transportation Use 36(60.0) | 51.5% |
| School Environment | 1 | School Cafeteria Use 38(63.3) | Classroom Rules Compliance 57(100) | Classroom Rules Compliance 57(95.0) | 86.1% |
| | 2 | Playground Use 34(56.7) | School Cafeteria Use 45(78.9) | Restroom Use 49(81.7) | 72.4% |
| | 3 | School Attendance 32(53.3) | Restroom Use 44(77.2) | School Cafeteria Use 45(75.0) | 68.5% |
| Peer Interaction | 1 | Birthday Party Participation 38(63.3) | Group Activity Participation 54(94.7) | Group Activity Participation 56(93.3) | 83.8% |
| | 2 | Board Game Participation 33(55.0) | Borrowing and Returning Items 49(86.0) | Borrowing and Returning Items 51(85.0) | 75.3% |
| | 3 | Meeting Friends by Appointment 26(43.3) | Meeting Friends by Appointment 35(61.4) | Meeting Friends by Appointment 45(75.0) | 59.9% |
| Classroom Activities | 1 | Arts and PE Participation 46(76.7) | Class Preparation 48(84.2) | Class Preparation 50(83.3) | 81.4% |
| | 2 | Science Class Participation 29(48.3) | Delivering School Announcement 37(64.9) | Delivering School Announcement 43(71.7) | 61.6% |
| | 3 | Presentation Participation 20(33.3) | Presentation Participation 28(49.1) | Keeping a Planner 33(55.0) | 45.8% |
| Career/ Vocational Training Situations | 1 | Coffee Making 29(48.3) | Item Sorting 36(63.2) | Job Interview 37(61.7) | 57.7% |
| | 2 | Item Assembly 25(41.7) | Item Assembly 31(54.4) | Item Sorting 36(60.0) | 52.0% |
| | 3 | Packaging Tasks 22(36.7) | Job Interview 31(54.4) | POS System Use 36(60.0) | 50.4% |

TABLEⅡ Comparison of Priority Rankings of Intervention Situations Across Groups

## IV. Conclusions

School-aged children are not merely passive recipients of clinical intervention but active, essential participants whose agency significantly influences the outcome of the intervention. The findings of this study provide empirical evidence that the perception of 'priority' regarding intervention situations exhibits a marked divergence among children, caregivers, and SLPs. While children consistently emphasize situations rooted in personal interests, social enjoyment, and authentic real-life participation, caregivers and SLPs tend to prioritize functional and pragmatic situations closely linked to academic performance, safety, and daily adaptation. This divergence suggests that a top-down approach, which exclusively reflects adult-centric perspectives, may overlook the motivational factors crucial for a child's sustained engagement. Therefore, these findings underscore that the selection of metaverse-based language intervention goals must not be a unilateral process. Instead, it should be established through shared decision-making, a collaborative framework that integrates the unique perspectives of children, the lived observations of caregivers, and the clinical expertise of professionals. By aligning these disparate priorities, developers and clinicians can create metaverse environments that are not only functionally robust for skill acquisition but also intrinsically motivating for the children, ultimately fostering a more holistic and effective therapeutic experience

## Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00218176) and the Soonchunhyang University Research Fund.

## V. References

[1] M. B. Hyeon and G. C. Yang, “Perspectives of speech-language pathologists on the utilization of virtual reality in speech pathology,” *Journal of Speech-Language & Hearing Disorders*, vol. 33, no. 2, pp. 125–134, 2024

[2] R. E. Owens, Jr., *Language Development: An Introduction*, 5th ed., Boston, MA: Pearson, 2008

[3] C. Baylor and M. Darling-White, “Achieving participation-focused intervention through shared decision making: Proposal of an age- and disorder-generic framework,” *American Journal of Speech-Language Pathology*, vol. 29, no. 3, pp. 1335–1360, 2020

# Describing Child Activities in Virtual Reality Using Multimodal Behavioral Signals and Large Language Models

T. Sokea[1], P. Samnang[1], L. Senghak[1], K. Chomyong[2], N. Yunyoung[3*]

[1]*Department of ICT Convergence, Soonchunhyang University, Asan 31538, Republic of Korea*
[2]*ICT Convergence Research Center, Soonchunhyang University, Asan 31538, Republic of Korea*
[3]*Emotional and Intelligent Child Care Convergence Center, Soonchunhyang University, Asan 31538, Republic of Korea*
**Corresponding Author: Yunyoung Nam. Email: ynam@sch.ac.kr*

*Abstract*—**Virtual reality (VR) learning environments capture high-frequency behavioral signals from head-mounted displays, including gaze targets, eye-point trajectories, object regions, head motion, hand-interaction states, facial action units, speech flags, and scene context. However, converting these multimodal telemetry streams into concise human-interpretable descriptions remains challenging. This study proposes a CSV-only large language model (LLM) framework for describing child activities in object-centric VR tasks. Meta Quest Pro telemetry is cleaned, resampled to 50 Hz, and segmented into non-overlapping 2-second windows. Longer human annotations are aligned to each model window using maximum temporal overlap on the video-time axis, and each window is summarized using gaze dwell and switching, gaze-in-object-region support, hand-activity energy, hold and gesture indicators, head-motion cues, facial-action-unit indices, speech flags, scene context, and object-name-enriched CSV rows. These structured summaries are provided to Qwen3.5-9B through Ollama using a constrained JSON output schema. Across 111 processed sessions and 3,441 generated 2-second windows, activity evaluation against human labels achieved 78.17% accuracy, 84.05% balanced accuracy, and 84.07% weighted F1-score. Target-object evaluation achieved 85.12% accuracy and 86.65% weighted F1-score. The results show that structured multimodal event extraction combined with evidence-grounded LLM prompting can support scalable and interpretable child-behavior reporting in VR learning and play environments.**

## I. Introduction

Virtual reality (VR) is increasingly adopted in education and child-centred training because it can deliver interactive tasks in a controlled setting while capturing detailed behavioral traces during learning. Compared to conventional observation, VR head-mounted displays can continuously record process-level signals such as eye gaze, head motion, facial expressions, and hand interactions. These signals are strongly linked to attention allocation and engagement, and they have been widely studied in VR for user interaction, education, and training applications [1–3]. In learning contexts, multimodal measures combining eye tracking and facial/physiological cues have shown value for understanding engagement and performance beyond outcome scores alone [4]. For children, multimodal learning analytics further highlights the benefit of integrating multiple behavioral channels to better interpret effort and learning-related behavior [5].

Despite this sensing capability, practical interpretation remains difficult. VR logs are typically high-frequency, high-dimensional time series (often hundreds of columns per frame), which makes manual review and annotation time-consuming and inconsistent. Many VR analytics studies focus on computing metrics (e.g., gaze duration, focus rates) or predicting performance/engagement [3], [6], but teachers and caregivers often need human-interpretable descriptions of what the child did and how they responded to instructions, such as "searched among toys," "focused on the requested object," or "attempted to pick it up." Producing such narratives at scale is challenging, particularly in tasks where behavior is object-driven (toy selection, manipulation) and must be interpreted in context (which objects were present, what the teacher requested, and whether the child complied).

Large language models (LLMs) provide a promising direction because they can generate fluent natural-language summaries. However, directly feeding raw numerical sensor logs to LLMs is unreliable: time-series signals contain noise and redundancy, and unconstrained generation can introduce unsupported statements (hallucinations) [7]. Recent work on aligning LLMs with sensor time-series emphasizes the importance of converting numeric sequences into semantically meaningful representations before language-based reasoning or description [8]. Surveys on LLMs for sensor and wearable behavior analysis similarly note that grounding, alignment, and domain constraints are essential for trustworthy outputs [9]. Therefore, for VR child activity description, an effective approach should first convert multimodal telemetry into compact, interpretable behavioral events (e.g., object-focused attention, gaze switching, interaction attempts), and then use an LLM to compose an evidence-grounded narrative aligned with task context.

In this study, we propose a VR-based framework to describe child activities in natural language from Meta Quest Pro–derived multimodal signals and VR scene context. Our setting

includes object-centric play and learning scenarios, where the VR application provides object identifiers and screen-space regions, and the child's behavior is captured through gaze targets, eye-point trajectories, head kinematics, hand movement/interaction states, and synchronized videos frames.

We segment sessions into short time windows. For each window, the system extracts structured multimodal events such as (i) sustained attention versus searching based on gaze dwell gaze switching, and gaze overlap with object regions; (ii) reach, touch, hold, movement, and placement cues inferred from hand motion, hand position, hold status, and gesture states; and (iii) focus or exploratory behaviour inferred from head motion dynamics. In addition, sampled video frames and low-level visual statistics provide complementary scene evidence for object presence and child-hand interaction.

These event summaries, together with an object and optional teacher instruction targets, are provided to a vision-language model/LLM to generated structured, evidence-grounded activity description. The human activity labels are used only for evaluation and reporting, not as input to the model during prediction.

The contributions of this work are:

- Multimodal event summarization for VR child behavior: We transform high-frequency gaze, eye-point, object-region, head-motion, hand-tracking, facial-action-unit, speech, and scene-context signals into interpretable 2-second window-level behavioral events.
- Evidence-grounded VLM/LLM activity narration: We generate child activity descriptions using a constrained output schema including primary activity, target object, eye-tracking class, hand-action class, destination object, compliance status, evidence, confidence, and narrative description to improve consistency and reduce hallucination risk [7].
- Object-centric instruction-following analysis: When teacher instruction targets are available, the framework links requested objects with gaze, hand-interaction, object-region, and video evidence to describe whether the child attended to, selected, manipulated, or placed the relevant object.
- Human-label-based evaluation: Human annotations are used as ground truth only after prediction to compute classification reports, confusion matrices, and consistency analyses for activity class, target object, eye focus, hand action, and destination object.

## II. Method

We implemented an end-to-end methodology for describing child activities in virtual reality from synchronized multimodal telemetry and video recordings. The pipeline consists of six stages: (1) discovery of synchronized CSV-video session pairs, (2) pre-processing and temporal resampling, (3) fixed-window segmentation, (4) extraction of interpretable multimodal behavioral evidence, (5) LLM-based generation of constrained window-level activity descriptions, and (6) merging, consistency checking, and evaluation.

### *1) Data Collection and Signals*

***A.*** *VR sessions and recordings*

Child VR sessions were collected in an object-centric learning/play scenario where the child interacts with virtual toys/objects and may receive instructions (e.g., to select a specific object). Each session produces synchronized telemetry logs exported as CSV and, when available, a recorded video stream for qualitative verification.

***B.*** *Multimodal signals and scene context*

The VR system logs multimodal channels at high temporal resolution. Table 1 summarizes the main signals used in this study and their behavioral meanings.

Table 1: Summary of multimodal VR signals and scene context used for child activity description

| Signal group | Example columns (from CSV) | Description and usage in this study |
|---|---|---|
| Gaze and eye features | l_eye_pnt_x, l_eye_pnt_y, r_eye_pnt_x, r_eye_pnt_y, eye_pnt_x, eye_pnt_y, l_eye_cls, r_eye_cls, gaz_obj | Captures visual attention and gaze behavior. Eye points provide gaze location, eye-closure values indicate blinking/eye openness, and gaz_obj provides the current gaze target object ID. Used to compute dwell time, gaze switching, fixation vs. search patterns, and attention stability. |
| Object regions (screen-space) | *_top_left_x, *_top_left_y, *_bottom_right_x, *_bottom_right_y (e.g., 1OA1_top_left_x, 1OA1_bottom_right_y) | Provides on-screen bounding regions for each VR object/toy. Used to geometrically verify whether the gaze point falls inside a specific object region (gaze-in-box ratio), improving robustness when gaz_obj is missing/noisy and supporting evidence grounding for LLM descriptions. |
| Head kinematics | head_roll, head_pitch, head_yaw, head_angul_vel | Represents head orientation and movement dynamics. Used to infer exploration/scanning behavior (high angular velocity) versus stable attention (low angular velocity), and to support interpretation of searching, turning, or focusing behaviors. |
| Facial action units (AUs) | brow_lowerer_l/r, lid_tightener_l/r, jaw_drop, lip_corner_puller_l/r, inner_brow_raiser_l/r, upper_lip_raiser_l/r | Continuous indicators of facial muscle activation. Used as supportive affect/effort cues (e.g., concentration/effort, surprise/arousal, positive engagement), which help explain behavior during searching, selection, or task difficulty moments. |

| | | |
|---|---|---|
| Hand behavior and interaction | l_hand_x, l_hand_y, r_hand_x, r_hand_y, l_hand_hld, r_hand_hld, l_hand_gest, r_hand_gest | Captures motor actions and interaction attempts. Hand positions are used to compute movement energy/velocity; hold and gesture flags indicate manipulation or interaction events. Combined with gaze to infer reach/pick/select behaviors and object manipulation attempts. |
| Speech and task context | is_speak, scene_act | Provides communication and scenario context. is_speak indicates verbal response or reaction during the task. scene_act encodes the task phase or scenario segment and is used to condition interpretation and segment-level narration. |

These multimodal signals provide complementary evidence to infer attention, search/exploration, interaction attempts, and engagement/affect, which are later summarized into window-level events and used to generate evidence-grounded activity descriptions.

***2) Preprocessing***

*A. Data cleaning and synchronization*

Each CSV is first sanitized by removing invalid values (NaN/Inf) and applying light smoothing where needed for stability (e.g., optional rolling median for jitter-prone channels). Timestamps (timer) are used to align all channels. If small gaps occur due to logging dropouts, short missing sequences are linearly interpolated for continuous signals (eye points, head angles, hand positions), while discrete variables (e.g., gaz_obj, flags) are forward filled within a short tolerance.

*B. Window segmentation*

The resampled telemetry was segmented into non-overlapping 2-second windows. With a target sampling rate of 50 Hz, each window was expected to contain approximately 100 samples, and windows with fewer than 80 valid samples were excluded. Although inference used 2-second windows, human activity labels were often annotated over longer intervals. Therefore, each 2-second window was aligned to the human label row having the maximum temporal overlap. When video timing (video_time_sec) was available, it was used as the primary label-alignment time axis because the manual labels were derived from the video timeline. If no matching label was found using video timing, the method fell back to timer-based alignment with an inferred offset.

For mixed human labels such as "Observing/Touching," the method computed the relative position of the 2-second window within the longer label interval and selected a phase-specific label action. This prevents the full longer label from being copied into both 2-second windows and supports finer interpretation of sequential behaviors.

***3) Multimodal Event Extraction***

For each 2-second window, high-dimensional telemetry was converted into compact interpretable event features. Gaze features included the dominant gaze object, object-name mapping, dwell ratio, gaze switch count, switch rate, and a rule-based search/fixation index. Object-region grounding was computed using a gaze-in-box ratio, defined as the proportion of valid gaze samples falling inside the screen-space bounding box of the dominant object. Hand features included left-hand, right-hand, and total hand activity energy computed from coordinate differences, as well as binary hold and gesture indicators. A reach/manipulation heuristic was activated when high hand activity coincided with stable object attention or when hold/gesture flags were detected.

Head-motion features were computed from head angular velocity when available, or from changes in roll, pitch, and yaw otherwise. Low head motion combined with stable gaze was used as a focus indicator, whereas high head motion was used as an exploration indicator. Facial action units were aggregated into concentration/effort, arousal/reactivity, and positive-engagement indices. The speech flag and parsed scene_act information were included as auxiliary context. The final event record for each window contained the time range, scene context, dominant object, gaze evidence, hand interaction evidence, head-motion evidence, facial-AU indices, speech flag, aligned human-label information, and optional instruction-compliance estimates. Table 2 summarizes the extracted events, their definitions, and how they are used for child activity description.

Table 2: Window-level multimodal event extraction for evidence-grounded narration

| Event category | Computed event / feature | Definition (per window (w)) | Behavioral meaning and usage |
|---|---|---|---|
| **Object-focused attention (gaze)** | Dominant gaze object | Most frequent gaz_obj observed within $(w)$. | Indicates the primary object of interest; used as the main candidate for "focused on / selecting / inspecting" behaviors. |
| | Dwell distribution | Time ratio on each object: $p(o) = \frac{gaz_{obj} = o}{t \in w}$ | Measures how long the child attends to each object; used to detect sustained attention and compare competing objects. |
| | Gaze switch count | Number of transitions between different gaz_obj values within $(w)$. | High switching suggests searching/scanning; low switching suggests stable attention or task commitment. |
| | Search vs. fixation index | Rule-based index derived from dwell concentration and switch rate. | Used to label behavior as "searching" (exploration) versus "fixating/focused" (stable attention) for narration. |

| | | | |
|---|---|---|---|
| **Geometry grounding (object regions)** | Gaze-in-box ratio | $inBox(o) = \frac{1}{N}\sum_{t\in w} 1\{x_t,\ y_t\ \in box_o(t)\}$, where $(x_t, y_t)$ is (eye_pnt_x, eye_pnt_y) and $box_o(t)$ is the recorded object region. | Strengthens evidence that the child truly looked at a specific object, especially when gaz_obj is uncertain, missing, or noisy. Used as grounding evidence for LLM outputs. |
| **Hand interaction & manipulation** | Hand activity energy | Velocity/acceleration magnitude of left and right hands (normalized), computed from l_hand_x/y, r_hand_x/y. | Indicates motor engagement; peaks suggest reaching, manipulating, or active interaction attempts. |
| | Interaction indicators | Presence of l_hand_hld, r_hand_hld, l_hand_gest, r_hand_gest within $(w)$. | Provides discrete evidence of holding/gesture actions; used to support "picked/held/selected/manipulated" descriptions. |
| | Reach/manipulation heuristic | Increased hand activity while gaze is stable on an object and/or hold/gesture activation. | Used to infer object interaction attempts (reach, grab, manipulate), especially when combined with dominant gaze evidence. |
| **Head motion cues** | Exploration indicator | Higher head_angul_vel within $(w)$. | Suggests scene scanning or searching; supports narration like "looking around" or "exploring options." |
| | Focus indicator | Low head motion combined with stable gaze (low switch count, high dwell). | Suggests concentration and sustained attention; supports narration like "focused on the object/task." |
| **Facial AUs & speech cues** | Concentration/ effort index | Compact AU summary derived from effort-related activations (e.g., brow lower, lid tightener). | Supports interpretation of task effort, difficulty, or focused engagement (used as supportive evidence, not as primary activity). |
| | Arousal/surprise index | AU summary from arousal-related activations (e.g., inner brow raiser, jaw drop). | Used to support descriptive context such as surprise/uncertainty or strong reaction during interaction. |
| | Positive engagement index | AU summary from positive affect cues (e.g., lip corner puller/smile-related). | Supports engagement interpretation (enjoyment/positive reaction) during successful interaction or play. |
| | Speech flag | is_speak within $(w)$ (binary). | Indicates possible verbal response, reading, or reaction; used to enrich descriptions when relevant. |
| **Event summary output** | Window event record | Structured record containing time range and scene_act; dominant object + dwell + switch + gaze-in-box; hand interaction features; head exploration/focus; AU indices + is_speak. | This compact event record is the primary input to the LLM, enabling evidence-grounded narration and reducing hallucination by constraining the model to measurable indicators. |

***4) LLM-Based Activity Description***

The current implementation generated window-level activity descriptions using Qwen3.5-9B through Ollama. The model was not provided with direct video frames. Instead, each prompt was constructed from CSV-derived behavioral evidence, including a structured event summary, a relevant object-ID map, temporally aligned human-label context, and a compact object-name-enriched CSV representation. This design allowed the LLM to generate descriptions from measurable VR telemetry while avoiding direct dependence on raw video interpretation.

To reduce prompt length and improve consistency, the CSV view was compressed before being passed to the LLM. Only task-relevant columns were retained, including timestamp, scene_act, gaz_obj, gaze-point coordinates, head orientation and angular velocity, hand coordinates, hold and gesture flags, speech flag, gaze-object name/category, and visible object units derived from object-region columns. Object IDs were converted into human-readable names using the object dictionary so that the LLM could refer to meaningful targets such as balls, boxes, animals, fruit, vehicles, or background regions rather than only numeric object codes.

*A. Prompt construction and grounding strategy*

For each 2-second window, the prompt contained three main sources of evidence. First, the structured event summary described the window time range, scene context, dominant gaze object, dwell ratio, gaze-switching behavior, gaze-in-box support, hand activity, hold/gesture indicators, head-motion cues, facial-action-unit indices, and speech flag. Second, the relevant object-ID map linked scene-specific object codes to human-readable object names and categories. Third, when a

human annotation overlapped the current window, the prompt included the aligned label information, including the full human action label, the phase-specific action for the current 2-second window, eye-tracking label, hand-tracking label, object involved, overlap ratio, and label phase information.

Because human annotations were commonly longer than the 2-second model windows, the prompt explicitly included temporal-overlap and phase information. This allowed the LLM to adapt a longer label, such as "Observing/Touching," to the specific part of the interval represented by the current 2-second window, rather than copying the full longer annotation into every window. The prompt also instructed the model to use only the provided evidence and to return "unknown" when the evidence was insufficient. This grounding strategy was intended to reduce unsupported generation and keep the output aligned with measurable gaze, hand, head-motion, object-region, and label-context evidence.

*B. Controlled output format*

To standardize the generated descriptions and support quantitative evaluation, the LLM was required to return a JSON object for each window. The output schema included primary_activity, target_object_id, target_object_name, compliance_status, evidence, confidence, and narrative. The primary_activity field was constrained to one of seven classes: searching, focusing, selecting, manipulating, responding, idle, or unknown. The target_object_id and target_object_name fields identified the most relevant object involved in the window, while the evidence field listed short statements supporting the prediction. The confidence score ranged from 0 to 1, and the narrative field provided a concise one- or two-sentence description of the child's behavior.

A rule-based fallback was applied when the LLM response failed or could not be parsed. This fallback used the extracted behavioral indicators, including gaze stability, dwell ratio, hand interaction, head motion, speech flag, and aligned human-label context, to produce a structured output in the same JSON-compatible format. Finally, adjacent windows with the same activity, target object, and compliance status were merged to form a smoother session-level summary for human review.

***5) Evaluation***

The framework was evaluated by comparing the LLM-generated window-level outputs with temporally aligned human annotations. Since the LLM produced 2-second descriptions while human labels were often longer, each window was matched to the human label segment with the maximum temporal overlap. Only windows with sufficient overlap were used for quantitative evaluation.

*A. Instruction-following and compliance performance*

For activity evaluation, the LLM output primary_activity was compared with the human child_action label after mapping both into the same classes: searching, focusing, selecting, manipulating, responding, idle, and unknown. When a human label contained multiple actions, such as "Observing/Touching," the phase of the 2-second window was used to select the most relevant sub-action. Performance was measured using accuracy, balanced accuracy, precision, recall, F1-score, and a confusion matrix. Error cases were also saved for detailed inspection.

*B. Narrative quality and evidence consistency*

For object evaluation, the LLM output target_object_name was compared with the human object_involved label. Object names were normalized before comparison to reduce differences caused by naming variations, such as "Basketballs," "Blue ball group," and "ball." The same metrics and confusion-matrix analysis were used to evaluate whether the generated description identified the correct object or scene element.

*C. Evidence consistency and error analysis*

Automated consistency checks were applied to examine whether the LLM outputs were supported by the extracted evidence. Target-object predictions were checked against dominant gaze, dwell ratio, gaze-in-box evidence, and aligned object labels. Selecting or manipulating predictions were checked against hand activity, hold, and gesture indicators, while searching and focusing predictions were checked against gaze switching, dwell, and head-motion cues. Error analysis was used to identify common mismatch cases, especially mixed-action labels, missing gaze-object values, multiple visible objects, and differences between broad human object labels and specific object-dictionary names.

## III. RESULTS AND DISCUSSION

***1) Data Distributions***

To validate that the multimodal signals used for event extraction are stable and physically plausible, we examine the eye and face groups before preprocessing, after outlier removal, and after interpolation-based reconstruction. Figures 1 and 2 illustrate representative time intervals, where the plots report the per-channel values against the session timer. Table 3 shows the number of session files used before preprocessing and after outlier removal with interpolation-based reconstruction.

***A.*** *Dataset coverage before and after preprocessing*

We report the number of session CSV files used for the distribution analysis before and after preprocessing. The raw dataset contains 169 session files, whereas the outlier-removed dataset contains 118 session files.

Table 3: Dataset coverage used for distribution analysis

| Dataset version | Normal | Abnormal | Total |
|---|---|---|---|
| Before preprocessing | 110 | 59 | 169 |
| After outlier removal & interpolation | 66 | 52 | 118 |

***B.*** *Eye-signal distributions (gaze points and eye-closure)*

The eye group includes left/right eye points, fused gaze point, and eye-closure indicators (e.g., l_eye_pnt_x/y, r_eye_pnt_x/y, eye_pnt_x/y, l_eye_cls, r_eye_cls). In the raw signals Fig. 1(a), the gaze-point trajectories show smooth trends in general, but they are affected by two common artifacts:

- Spikes and discontinuities: brief abrupt deviations in gaze coordinates and/or closure values that are inconsistent with typical eye motion at the sampling rate.
- Dropout-induced gaps: missing samples appearing as broken segments, often occurring when tracking

confidence decreases (e.g., partial occlusion, headset fit changes, or rapid head movement).

Figure 1 illustrates representative eye-signal distributions before and after pre-processing.

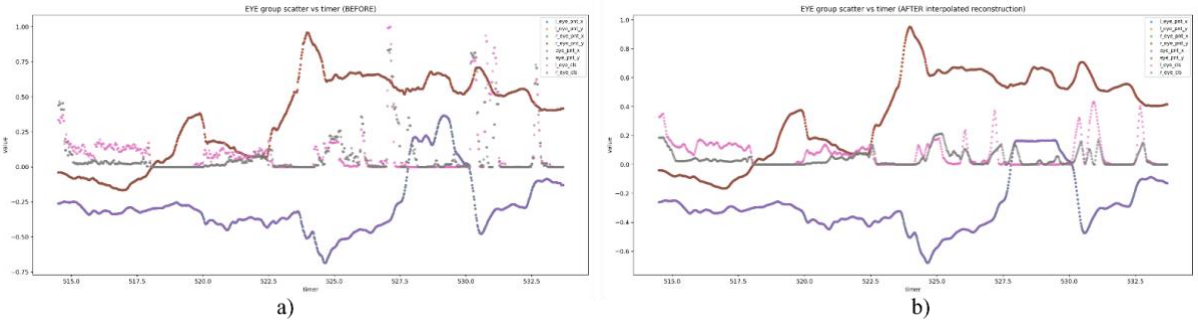



Figure 1: Eye-signal distributions before and after pre-processing: (a) raw signals before pre-processing and (b) signals after outlier removal and interpolation-based reconstruction. The reconstruction improves continuity of gaze trajectories and stabilizes eye-related signals for window-level event extraction.

After outlier removal, extreme spikes are suppressed and the remaining gaze trajectories better reflect plausible motion. However, removing outlier rows introduces temporal fragmentation, producing visible gaps that can interrupt short-window statistics (e.g., gaze velocity, saccade-like changes, dwell stability). Finally, after interpolation-based reconstruction Fig. 1(b), these short gaps are filled for continuous variables, restoring a near-continuous gaze trajectory while preserving the overall trends. This improves the reliability of gaze-derived events (dwell, switching, fixation vs. search), because window-level computations are less sensitive to missing samples.

Importantly, the eye-closure signals (l_eye_cls, r_eye_cls) remain sparse and burst-like across all stages, which is expected because blinking is intermittent. The preprocessing primarily reduces unrealistic isolated peaks and stabilizes the time alignment of closure bursts with gaze trajectories.

### *C. Face-signal distributions (facial action units)*

The face group consists of facial AUs representing brow, eyelid, jaw, and lip movements. In the raw signals Fig. 2(a), AU channels exhibit a dense near-zero baseline with intermittent increases, reflecting that many facial muscles are inactive for long intervals while specific expressions activate briefly. Compared to gaze, AU channels typically show higher susceptibility to estimation noise, visible as isolated spikes or short-lived bursts inconsistent with surrounding context. Figure 2 shows representative facial-action-unit distributions before and after preprocessing.

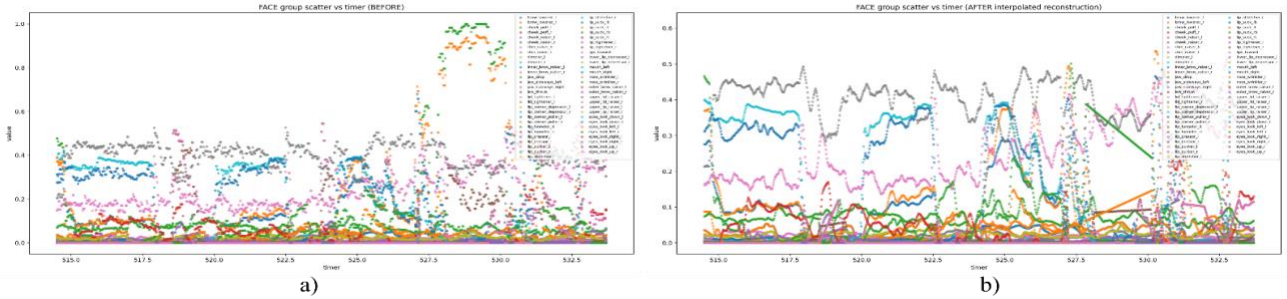



Figure 2: Face-group AU signals over time: (a) before preprocessing, and (b) after outlier removal (rows dropped) and after interpolation-based reconstruction. Outlier suppression reduces implausible spikes, and reconstruction restores temporal continuity for stable AU aggregation.

Outlier removal reduces these extreme excursions and improves the plausibility of AU magnitudes but similarly introduces gaps because entire rows are removed. After interpolation-based reconstruction Fig. 2(b), short gaps in continuous AU channels are filled, yielding smoother temporal evolution. This is essential for computing compact AU indices (e.g., effort/concentration, arousal/surprise, positive engagement) because these indices rely on windowed aggregation and should not be dominated by single-frame artifacts.

The final merged output contained 111 processed VR sessions, including 66 Normal sessions and 45 Abnormal sessions. After preprocessing, resampling to 50 Hz, and non-overlapping 2-second segmentation, the system generated 4357 window-level event records and 4357 LLM description records. The window distribution was 2621 Normal windows and 1736 Abnormal windows. For quantitative comparison with human annotations, only windows with a label-overlap ratio of at least 0.50 were retained, resulting in 2691 comparable windows. The final processing and evaluation coverage is summarized in Table 4.

Table 4: Dataset and evaluation coverage after preprocessing and 2-second segmentation

| Item | Value | Description |
|---|---|---|
| Processed sessions | 111 | Final merged sessions used for description generation |
| Normal / Abnormal sessions | 66 / 45 | Session-level group distribution |
| 2-second event windows | 4,357 | Total event records |

| | | |
|---|---|---|
| Normal / Abnormal windows | 2,621 / 1,736 | Window-level distribution |
| Activity-comparison windows | 2,691 | Windows with label-overlap ratio >= 0.50 |
| Activity metric windows | 2,662 | After removing unknown classes |
| Object metric windows | 2,251 | After removing unknown object classes |

### *D. Activity-Class Evaluation*

Table 5 summarizes the overall comparison between LLM-generated primary_activity and mapped human activity labels, and Table 6 reports class-specific performance. The model achieved 78.17% accuracy and 84.05% balanced accuracy over 2662 metric windows. The weighted F1-score was 84.07%, indicating strong overall agreement after considering class frequency. The macro F1-score was lower (51.62%), reflecting class imbalance and difficulty for rare or ambiguous behaviors.

Table 5: Overall activity-class evaluation metrics

| Metric | Value |
|---|---|
| Metric windows | 2,662 |
| Accuracy | 78.17% |
| Balanced accuracy | 84.05% |
| Macro precision | 53.31% |
| Macro recall | 70.04% |
| Macro F1 | 51.62% |
| Weighted precision | 94.04% |
| Weighted recall | 78.17% |
| Weighted F1 | 84.07% |

Table 6: Per-class activity performance for classes with human-label support.

| Activity class | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| searching | 19.00% | 97.44% | 31.80% | 39 |
| focusing | 97.04% | 76.85% | 85.77% | 1,577 |
| selecting | 90.88% | 64.01% | 75.11% | 389 |
| manipulating | 99.44% | 88.18% | 93.47% | 609 |
| idle | 13.47% | 93.75% | 23.56% | 48 |

The corresponding activity-confusion pattern is visualized in Fig. 3.

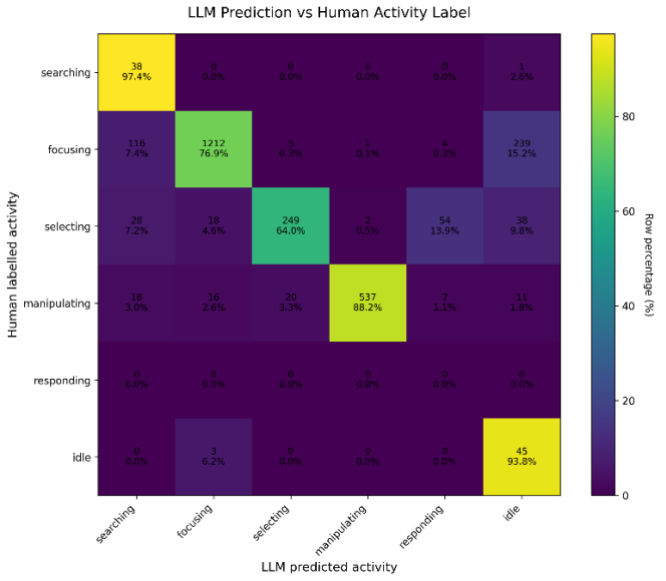



Figure 3: Activity-class confusion matrix for LLM-generated primary activity prediction. The matrix compares LLM-generated primary_activity labels with mapped human activity labels. Each cell reports the number of windows and the row-normalized percentage for each human activity class.

The activity confusion matrix indicates that manipulating, focusing, searching, and idle were generally identifiable from the structured evidence. Searching and idle obtained high recall, while selecting was more difficult because selecting behavior can share evidence with focusing, responding, or manipulation. Focusing errors frequently occurred when stable looking, weak movement, or exploratory head motion made the distinction between sustained attention and non-interactive behavior ambiguous in short 2-second windows.

### *E. Target-Object Evaluation*

Target-object evaluation compared the LLM-generated target_object_name with the normalized human object_involved label. As shown in Table 7, object prediction achieved 85.12% accuracy, 80.21% balanced accuracy, and 86.65% weighted F1-score over 2251 metric windows. These results show that object-ID mapping, gaze-object signals, visible object units, and label-grounded context helped the model identify the relevant VR object in most windows. Object-level performance for high-support categories is reported in Table 8, and the corresponding object-level confusion matrix is shown in Fig. 4.

Table 7: Overall target-object evaluation metrics.

| Metric | Value |
|---|---|
| Metric windows | 2,251 |
| Accuracy | 85.12% |
| Balanced accuracy | 80.21% |
| Macro precision | 89.71% |
| Macro recall | 74.56% |
| Macro F1 | 79.76% |
| Weighted precision | 90.81% |
| Weighted recall | 85.12% |
| Weighted F1 | 86.65% |

Table 8: Target-object performance for object classes with support >= 20.

| Object class | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| ball | 92.95% | 95.24% | 94.08% | 609 |
| teacher | 97.65% | 80.21% | 88.08% | 571 |
| animal | 89.66% | 83.42% | 86.43% | 187 |
| container | 46.60% | 82.76% | 59.63% | 174 |
| toy | 100.00% | 90.38% | 94.95% | 104 |
| fruit | 94.81% | 91.25% | 92.99% | 80 |
| vehicle | 98.15% | 91.38% | 94.64% | 58 |
| fish | 89.09% | 92.45% | 90.74% | 53 |
| clear bins | 100.00% | 81.40% | 89.74% | 43 |
| background | 29.35% | 79.41% | 42.86% | 34 |
| human figurine | 100.00% | 78.57% | 88.00% | 28 |
| floor | 100.00% | 52.00% | 68.42% | 25 |

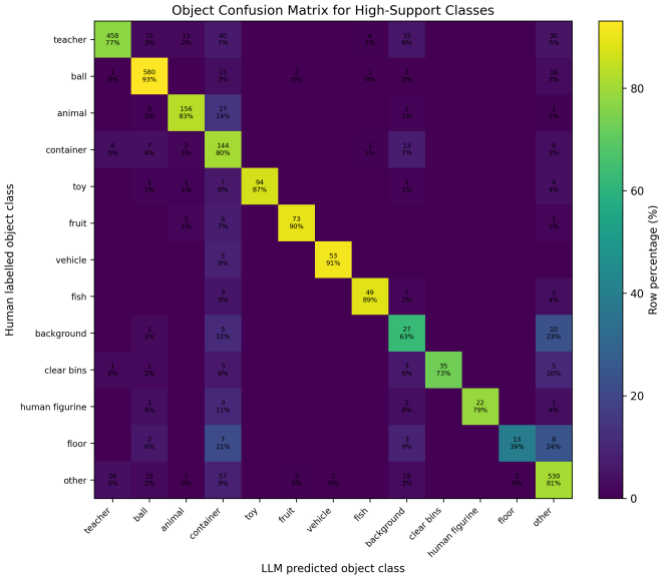



Figure 4: Readable object confusion matrix summarizing high-support object classes and grouping low-support classes as other.

Object prediction was strongest for frequent and semantically clear categories such as ball, teacher, toy, vehicle, fruit, fish, and animal. Lower performance for background and container-like categories reflects differences in annotation granularity. For example, human labels may use broad names such as tray or container, whereas the object dictionary may contain more specific objects such as transparent box, clear bin, or scene-region identifiers.

### *F.* Evidence Consistency and Error Analysis

Automated evidence-consistency checks were computed over all 4357 generated windows. The overall all-pass rate was 62.98%. Individual support rates were higher: object support was 83.98%, interaction support was 86.34%, search support was 92.79%, and focus support was 95.62%. The lower all-pass rate is expected because a generated window must satisfy all consistency rules simultaneously. Table 9 summarizes the automated evidence-consistency checks across all generated windows.

Table 9: Automated evidence-consistency rates over all generated windows.

| Consistency check | Rate |
|---|---|
| All checks passed | 62.98% |
| Object supported | 83.98% |
| Interaction supported | 86.34% |
| Search supported | 92.79% |
| Focus supported | 95.62% |

The activity error file contained 581 mismatched activity windows after filtering, and the object error file contained 335 mismatched object windows. Activity errors often occurred when a longer human label contained multiple actions, such as observing followed by touching, while the 2-second window represented only one phase. Object errors often occurred when several objects were visible or when the human annotation used a broader object name than the object dictionary.

### *G.* Discussion

The results support the feasibility of using structured multimodal telemetry and evidence-grounded LLM prompting to describe child behavior in object-centric VR tasks. The weighted activity F1-score of 84.07% indicates that generated primary activity labels were generally consistent with human annotations when class frequency was considered. Stronger performance for manipulating and focusing suggests that gaze stability, object dwell, hand activity, hold/gesture flags, and head-motion features provide useful evidence for distinguishing active object interaction from visual attention.

Target-object prediction showed stronger overall performance than activity prediction, with 85.12% accuracy and 86.65% weighted F1-score. This suggests that object-ID mapping and object-name-enriched CSV rows provide useful grounding for the narrative. However, object-level evaluation also revealed a granularity issue: human labels sometimes use broad terms such as ball, teacher, tray, or container, whereas the system may output a more specific object or region name.

Several limitations remain. First, the current implementation is CSV-only and does not provide video frames directly to the LLM. Second, human annotations are longer than the inference windows, so disagreement may occur when a human label covers multiple actions. Third, object-level performance depends on the consistency of object naming and normalization. Future work should incorporate direct video-frame evidence, refine object taxonomy mapping, and collect finer temporal annotations for more precise evaluation.

## IV. Conclusion

This study presented a CSV-only framework for describing child activities in virtual reality using multimodal behavioral telemetry and a large language model. The method preprocesses and resamples Meta Quest Pro CSV logs, segments sessions into 2-second windows at 50 Hz, aligns longer human annotations to each window using video-time overlap, extracts interpretable behavioral evidence, and uses a constrained Qwen/Ollama prompt to generate structured activity descriptions. The extracted evidence includes gaze dwell and switching, gaze-in-object-region support, hand activity and interaction indicators, head-motion cues, facial-action-unit indices, speech flags, scene context, and object-name-enriched CSV rows.

The final evaluation on 111 sessions and 4357 generated windows showed that the proposed method achieved 78.17% activity accuracy and 84.07% weighted activity F1-score against human labels. For target-object prediction, the method achieved 85.12% accuracy and 86.65% weighted F1-score. These findings demonstrate that structured multimodal feature extraction combined with evidence-grounded LLM prompting can generate coherent and quantitatively evaluable descriptions of child behavior in VR learning and play environments.

Future work will extend the framework by incorporating direct video-frame evidence, refining object-category normalization, improving handling of rare activity classes, and collecting finer temporal annotations. The proposed framework can support scalable behavioral reporting, teacher review, child-care monitoring, and dataset creation for downstream VR activity-recognition models.

## Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No RS-2023-00218176).

## REFERENCES

[1] I. B. Adhanom, S. Lee, A. K. Dey, and S. E. Hudson, "Eye Tracking in Virtual Reality: A Broad Review of Applications and Challenges," ACM Comput. Surv., 2023.

[2] R. Shadiev, Y. Wang, and M. Huang, "A review study on eye-tracking technology usage in immersive virtual reality learning environments," Computers & Education, vol. 196, 2023.

[3] M. Mikhailenko, M. Womser-Hacker, and M. Bartoschek, "Eye-Tracking in Immersive Virtual Reality for Education: A Review of the Current Progress and Applications," Frontiers in Education, 2022.

[4] I. Dubovi, A. Levy, and Y. Dagan, "Cognitive and emotional engagement while learning with VR: A multimodal approach using eye tracking and facial expression analysis," Computers & Education, 2022.

[5] (Systematic review) "Multimodal Learning Analytics research with young children: A systematic review," 2020.

[6] Y. Lin et al., "A method for evaluating learning concentration in head-mounted virtual interaction," Virtual Reality, 2023.

[7] L. Huang et al., "A Survey on Hallucination in Large Language Models," ACM Computing Surveys, 2025.

[8] Z. Li et al., "SensorLLM: Aligning Large Language Models with Motion Sensor Time-Series for Human Activity Recognition," in Proc. EMNLP, 2025.

[9] E. Ferrara, "Large Language Models for Wearable Sensor-Based Human Activity Recognition and Behavior Modeling: Trends and Challenges," Sensors, vol. 24, no. 15, 2024.

# Training-Free Parent-Child Interaction Detection from Multi-Camera Surveillance Using Vision-Language Models: A Spatiotemporal Collage Approach with LLM-as-Judge Evaluation and Semantic Retrieval

*Samnang Pheng[1], Sokea Teng[2], Senghak Lim[1], Yaena Ha[2], Chomyong Kim[2], and Yunyoung Nam[3]**

*[1]Department of Computer Science and Engineering, Soonchunhyang University, Asan 31538, Republic of Korea*
*[2]ICT Convergence Research Center, Soonchunhyasng University, Asan 31538, Republic of Korea*
*[3]Emotional and Intelligence Child Care System Convergence Research Center, Soonchunhyang University, Asan 31538, Republic of Korea*
**Corresponding Author: Yunyoung Nam. Email: ynam@sch.ac.kr*

**Abstract—**Assessing parent-child interaction (PCI) quality is central to developmental psychology research, yet traditional observational coding is labor-intensive, requires specialist training, and cannot scale to large video archives. We present a training-free, criterion-grounded framework that automates PCI detection from multi-camera closed-circuit television (CCTV) footage using a Vision-Language Model (VLM). Our approach constructs a 3×3 spatiotemporal collage: three camera angles across three consecutive one-second frames and submits it to Qwen3-VL-8B-Instruct with a structured prompt encoding five expert-defined interaction criteria inspired by the Dyadic Parent-Child Interaction Coding System (DPICS). Across six naturalistic play sessions totaling 1,304 three-second collages, our model achieves a macro-averaged G-mean of 0.77, accuracy of 79.75%, and F1 of 87.12%, substantially outperforming five competing baselines (Qwen3-VL-8B-Thinking, Gemma-3-4B, Aya-Vision-8B, InternVL3-8B, and Nemotron-8B), none of which exceeds a G-mean of 0.52, and two of which (Gemma-3-4B and Aya-Vision-8B) fully collapse to degenerate always-Yes prediction (G-mean = 0.00). We further introduce two complementary modules: (1) an LLM-as-Judge pipeline using gpt-oss-120b (Cerebras) to evaluate reasoning quality on a 1–10 scale against expert behavioral annotations, and (2) an Analysis Retrieval system built on ChromaDB with Qwen3-Embedding-4B and Qwen3-Reranker-4B that enables semantic search over 1,304 per-collage observations and video-level narrative summaries generated via map-reduce using Gemini 3.0 flash. Our results demonstrate that carefully prompted VLMs can match expert-level binary classification without any task-specific fine-tuning, and that the complete pipeline from raw video to searchable interaction archive is viable in a research deployment context.

**Keywords:** parent-child interaction; vision-language models; spatiotemporal collage; training-free classification; multi-camera surveillance; LLM-as-judge; semantic retrieval; map-reduce summarization; developmental psychology

## I. INTRODUCTION

The quality of early parent-child interaction is one of the strongest predictors of children's cognitive, linguistic, and socio-emotional development [1]-[3]. Clinical and research assessments of PCI traditionally rely on manual observational coding systems, most prominently the Dyadic Parent-Child Interaction Coding System (DPICS) [4][5], in which trained raters watch recorded play sessions and label micro-behaviors such as mutual gaze, joint attention, physical contact, and verbal engagement. While these systems achieve strong inter-rater reliability, they impose severe scalability constraints: a single one-hour session may require several hours of expert review, making large-scale deployment practically infeasible [6].

Automated video analysis offers an alternative path. Early computer-vision approaches relied on hand-crafted features, pose estimation, and gaze estimation pipelines [7]. The emergence of large pre-trained Vision-Language Models (VLMs), capable of grounded visual reasoning from natural-language instructions, opens the possibility of training-free behavioral coding without task-specific training data [8]-[10]. However, applying VLMs to behavioral observation raises several challenges: (i) naturalistic play sessions involve subtle, context-dependent cues requiring spatial reasoning across multiple camera viewpoints; (ii) interaction rate is typically high (≈85% of time windows in our corpus), creating severe class imbalance; and (iii) model outputs must be interpretable enough to audit against expert annotations.

This paper makes the following contributions:

**(1) Spatiotemporal collage prompting:** We introduce a 3×3 grid placing three camera perspectives in rows and three consecutive one-second frames in columns, embedding spatial and temporal positional information directly into the image to exploit VLM reasoning capabilities without fine-tuning.

**(2) Expert-grounded in-context PCI classification:** We design a structured five-criterion prompt operationalizing DPICS-inspired criteria (mutual gaze, parent- and child-directed attention, physical contact, shared attention), including a multi-step evidence quality gate requiring multi-camera or multi-frame corroboration.

**(3) Comprehensive multi-model benchmark:** We compare against five competing VLMs (Qwen3-VL-8B-Thinking, Gemma-3-4B, Aya-Vision-8B, InternVL3-8B, and Nemotron-8B) across six naturalistic play-session datasets (1,304 collages total), reporting G-mean as the primary metric to account for class imbalance.

**(4) LLM-as-Judge reasoning evaluation:** We develop an automated pipeline that assesses the factual consistency of per-collage model reasoning against structured expert behavioral descriptions using a large LLM as judge on a 1–10 scale.

**(5) Analysis Retrieval system:** We build a semantic vector database over all model-generated observations using Qwen3-Embedding-4B for indexing and Qwen3-Reranker-4B for re-ranking, enabling natural-language retrieval alongside video-level narrative summaries via map-reduce.

## II. RELATED WORK

### A. Automated Behavioral Coding of PCI

Manually coded systems such as DPICS [4][5] have been the gold standard in developmental psychology for decades. Karaca et al. [11] provides the most comprehensive recent survey of automated methods for nonverbal behavior analysis in PCI, identifying a critical shortage of public datasets for naturalistic

settings. Early automation efforts applied NLP to transcript-based DPICS classification [12] and pose estimation to social touch and gaze direction [7][13]. More recent work quantifies parenting quality from home-visit recordings without specialized hardware [14]. The state of the art is represented by HARMONI [15], which applies 3D mesh reconstruction to 500 hours of naturalistic home video and achieves 63–80% consistency with human annotators. Our work extends this line by replacing specialized computer vision pipelines with a single in-context prompted VLM call.

### *B.* Vision-Language Models for Video Understanding

GPT-4V [8] and LLaVA [9] pioneered instruction-following VLMs based on CLIP-style vision encoders coupled to large language models. Qwen-VL [10], Qwen2-VL [16], and Qwen3-VL [17] introduced dynamic resolution encoding and patch positional embeddings. InternVL [18] scales vision foundation models alongside language models. For video-specific understanding, Video-ChatGPT [19], VideoLLaMA 2 [20], VideoChat [21], and TimeChat [22] frame temporal reasoning as instruction-following tasks. Video-MME [23] documents consistent performance drops in the long-video regime that motivate our temporal chunking approach.

Three concurrent works apply multimodal LLMs directly to PCI analysis. Shi et al. [24] evaluate MLLMs on 26 PCI videos annotated by speech-language pathologists, finding that current MLLMs struggle with child-initiated eye contact, the gap our multi-view corroboration addresses. Shi and Choo [25] achieve 85% accuracy on perceptual cue extraction in a two-stage pipeline, and Shi and Choo [26] report >75% average precision on expert judgment simulation, broadly consistent with our macro-averaged precision of 94.40% and accuracy of 79.75% across six sessions, though task formulations differ. CollagePrompt [27] introduced and benchmarked tiling multiple images into a single prompt, demonstrating that spatial layout substantially affects recognition accuracy; our 3×3 collage directly instantiates this paradigm.

### *C.* Multi-Camera Analysis and Joint Attention

Li et al. [28] tackle automated joint attention and mutual gaze detection in free-play PCI, demonstrating that naturalistic free-play settings require multi-view or multi-cue fusion to reliably localize gaze targets. Beyan et al. [29] survey automated dyadic co-located interaction analysis using nonverbal cues, identifying lack of annotated naturalistic datasets as the primary research bottleneck. Our collage approach avoids explicit camera calibration by representing multi-view information directly as a spatial image layout, leveraging VLM visual reasoning to integrate information across camera perspectives.

### *D.* LLM-as-Judge Evaluation

Zheng et al. [30] proposed MT-Bench and the LLM-as-judge paradigm, where a large language model evaluates model output quality relative to reference answers. We adapt this paradigm to behavioral science: our judge compares VLM-generated model reasoning to expert behavioral annotation descriptions using a two-level rubric assessing both interaction-presence agreement and criterion-level alignment.

### *E.* Semantic Retrieval and Long-Context Summarization

Dense retrieval combines learned text embeddings with approximate nearest-neighbor search to retrieve relevant documents from large corpora using cosine similarity search [31]. Qwen3-Embedding-4B and Qwen3-Reranker-4B provide state-of-the-art dense retrieval and cross-encoder reranking pipelines [32]. Unlike retrieval-augmented generation (RAG), which couples retrieval with a generative model to synthesize new answers, our analysis retrieval system performs pure semantic search: per-collage observations are encoded as text embeddings, stored in ChromaDB, and retrieved by nearest-neighbor search followed by cross-encoder reranking, with no generation step. The LLM×MapReduce framework [33] formalizes map-reduce processing: documents are chunked, processed independently (Map), and aggregated (Reduce), the pattern we use to pre-compute per-session video narratives from per-collage observations before indexing.

### *F.* Evaluation Metrics for Imbalanced Classification

Standard accuracy is misleading for imbalanced datasets. The geometric mean G-mean = $\sqrt{(TPR \times TNR)}$ rewards balanced performance and penalizes degenerate classifiers by returning 0 whenever either TPR or TNR is 0 [34]. Luque et al. [35] confirm G-mean is among the most stable metrics under class imbalance. We report G-mean as the primary metric, alongside accuracy, precision, recall, and F1.

## III. DATASET

### *A.* Data Collection

Our corpus consists of six naturalistic parent-child play sessions recorded in a research playroom using a fixed multi-camera CCTV system. Each session captures a single parent-child dyad engaged in free play with age-appropriate toys. Three calibrated cameras provide complementary viewing angles: an overhead perspective, a frontal wide-angle view, and a close-up angle focusing on the interaction space. Sessions are identified as S-006, S-011, S-012, S-022, S-026, and S-037.

### *B.* Collage Generation

Raw video footage is segmented into non-overlapping three-second windows, one frame per second. For each window, frames at t+0 s, t+1 s, and t+2 s are extracted from all three cameras, producing a 3×3 image grid (Fig. 1). Each row corresponds to one camera angle; each column corresponds to one time step. This 3×3 collage is the sole visual input to the VLM. The final dataset contains 1,304 collages across six sessions (Table I).



Figure 1. Example 3×3 spatiotemporal collage. Rows correspond to three camera angles (overhead, frontal wide-angle, and close-up); columns correspond to one-second frames at t+0 s, t+1 s, and t+2 s. This grid is the sole visual input to the VLM.

Table I. Dataset statistics per session. Overall class imbalance: 84.7% positive.

| Session | Collages | Yes | No | Yes % |
|---|---|---|---|---|
| S-006 | 189 | 172 | 17 | 91.0% |
| S-011 | 198 | 170 | 28 | 85.9% |
| S-012 | 197 | 177 | 20 | 89.8% |
| S-022 | 295 | 233 | 62 | 79.0% |
| S-026 | 194 | 140 | 54 | 72.2% |
| S-037 | 231 | 212 | 19 | 91.8% |
| **Total** | **1,304** | **1,104** | **200** | **84.7%** |

### C. Ground Truth Annotation

Ground truth binary labels (Interaction: Yes/No) were assigned by a trained researcher using a structured annotation protocol based on five interaction criteria adapted from DPICS:

- Criterion 1. Parent-directed gaze: Parent's eye/head orientation shows attention directed at the child's face or head, sustained across cameras or timepoints.
- Criterion 2. Child-directed gaze: Child clearly looks at the parent; child must show awareness or response if this is the sole criterion.
- Criterion 3. Mutual gaze: Both clearly look at each other's face/head at the same timepoint.
- Criterion 4. Physical contact: Visible touching, hand-holding, or direct object handoff; not satisfied by mere arm extension.
- Criterion 5. Shared attention: Both clearly focused on the identical specific object simultaneously; shared location or general area is insufficient.

A collage is labeled Yes if at least one criterion is clearly met and corroborated. Ground truth descriptions were recorded separately to support LLM-as-Judge evaluation (Section IV-F).

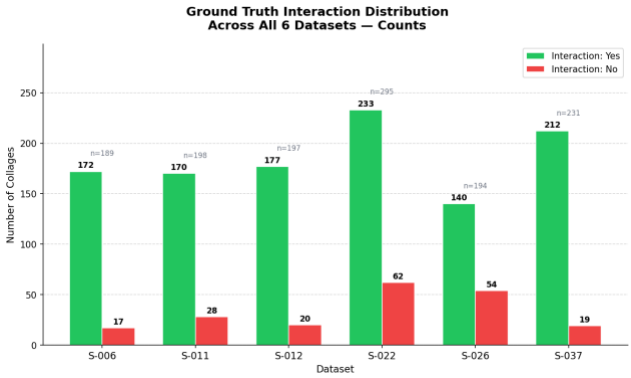



Figure 2. Ground truth label counts per session. Interaction is dominant in all sessions; S-026 shows the most balanced distribution (72.2% positive).

## IV. Proposed Methods

### A. System Architecture

The architecture diagram illustrates the full pipeline comprising four stages: (1) spatiotemporal collage generation, (2) VLM inference with structured prompting, (3) Result generation with summarization, and (4) Analysis Retrieval over the resulting observation archive.

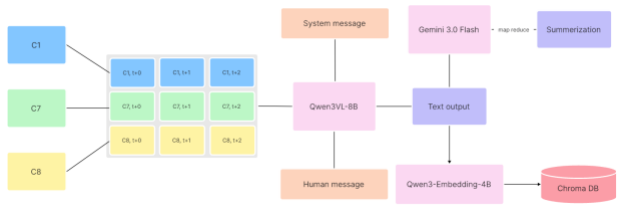



Figure 3. System architecture. Multi-camera video is segmented into 3-second windows, converted into 3×3 spatiotemporal collages, and analyzed by Qwen3-VL-8B-Instruct. All the structured output feeds into Gemini 3.0 flash for map reduce summerization and a chroma db for semantic retrieval system.

### B. Vision-Language Model: Qwen3-VL-8B-Instruct

We use Qwen3-VL-8B-Instruct [17], an 8-billion-parameter multimodal model built on the Qwen3 language backbone [36] that extends Qwen2-VL [16] with dynamic resolution encoding, interleaved MRoPE for spatiotemporal modeling, and DeepStack vision-language alignment. We use the model in training-free mode; no task-specific fine-tuning is applied. Inference runs on a single NVIDIA GPU (24 GB VRAM) with bfloat16 precision at approximately 40–60 collages per hour.

### C. Prompt Design

The prompt is divided into two components: a system prompt defining the collage structure and grid layout; and a user prompt providing identity information about the dyad, interaction criteria and instructing the model to follow a four-step decision process.

***Identity Extraction.***

Before the main interaction analysis, a brief identity-extraction step asks the model to describe each person's complete outfit from a reference frame. This description is embedded under PERSON IDENTITIES, preventing the model from confusing parent and child when their positions overlap between camera angles.

***Interaction Criteria and Decision Process.***

The user prompt specifies the five criteria with detailed behavioral operationalizations, exclusion rules (e.g., ALWAYS mark No for these), and ten worked calibration examples (A–J). The decision process has four steps: (1) VERIFY PRESENCE, confirm both persons are visible; (2) PARALLEL PLAY CHECK, reject cases where each person is independently absorbed; (3) CHECK EACH CRITERION, evaluate each criterion with camera and timepoint citations; (4) CONSISTENCY CHECK, ensure criteria, confidence, and label are internally consistent.

***Evidence Quality Gate.***

A critical element of our prompt is the corroboration requirement: criteria 1, 2, and 3 require either (a) two different cameras confirming the criterion at the same timepoint, or (b) the same camera confirming it at two consecutive timepoints. This gate substantially reduces false positives by requiring evidence beyond a single ambiguous cue.

***Output Format.***

The model produces structured plain-text and JSON outputs containing: Collage ID, Time Range, Interaction (Yes/No), Confidence (high/medium/low), Evidence section (visual evidence, contact check, model reasoning, human-intuition summary), Ambiguity note, and Criteria Met list.

### D. Comparison Baselines

We compare against five baselines: Qwen3-VL-8B-Thinking (same architecture with chain-of-thought reasoning enabled); Gemma-3-4B [37] (Google's 4B instruction-tuned VLM); Nemotron-8B (NVIDIA Llama-3.1-Nemotron-Nano-VL-8B-V1); InternVL3-8B [18] (OpenGVLab's 8B VLM); and Aya-Vision-8B (Cohere's multilingual VLM). All comparison VLMs receive the identical five-criterion prompt without model-specific optimization.

### E. Evaluation Metrics

Given the class imbalance (≈85% positive), we adopt G-mean as the primary metric: G-mean = $\sqrt{(TPR \times TNR)}$. G-mean is bounded in [0,1] and equals 0 whenever either sensitivity or specificity is zero, directly penalizing degenerate classifiers. We additionally report accuracy, precision, recall, and F1.

### F. LLM-as-Judge Reasoning Evaluation

Beyond binary classification accuracy, we assess the quality of model reasoning using an automated LLM judge. For each collage, the judge receives: (a) a ground truth behavioral description from the expert annotator, and (b) the model's visual evidence or model

reasoning section. The judge (gpt-oss-120b via Cerebras API) scores factual consistency on a 1–10 scale using a two-level rubric. Level 1: interaction presence agreement (correct TP/TN receives baseline 5+; incorrect FP/FN receives 1–3); Level 2: criterion alignment (same criterion adds 2–3 points).

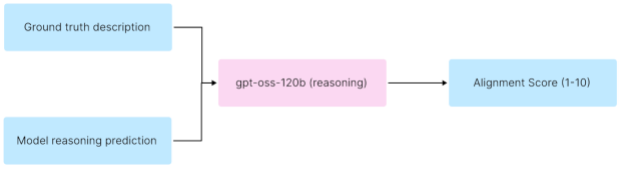



Figure 4. LLM-as-Judge pipeline. The judge compares the model's reasoning against expert behavioral annotations, producing a 1–10 factual consistency score.

### G. Analysis Retrieval System

To make the 1,304-collage observation archive accessible for downstream research queries, we build a semantic vector database (Fig. 5). Following the LLM×MapReduce paradigm [33], each session's collage observations are chunked into groups of 25 and individually summarized via MAP LLM calls (Gemini 3.0 flash). Chunk summaries are then combined by a REDUCE call producing a 3–5 paragraph narrative summary covering interaction quality, temporal evolution, notable periods, and dominant interaction modalities. Per-collage documents, combining the model's human-intuition summary and a visual evidence excerpt (up to 600 characters), together with video-level narrative summaries are encoded using Qwen3-Embedding-4B [32] (embedding dimension: 2,560; max context: 32k tokens). Queries use an asymmetric prompt_name="query" encoding path. Embeddings are stored in ChromaDB using cosine similarity for retrieval (≈1,310 indexed documents across six sessions). Query results are reranked by Qwen3-Reranker-4B using generative CausalLM scoring.

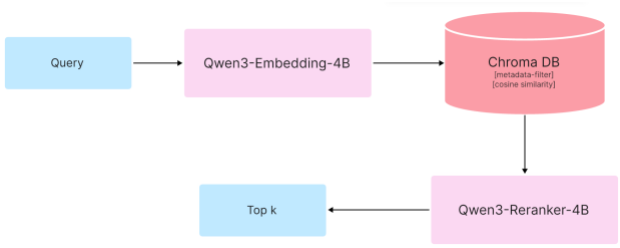



Figure 5. Analysis Retrieval pipeline. Per-collage observations and session-level narratives are embedded with Qwen3-Embedding-4B, stored in ChromaDB, and retrieved with Qwen3-Reranker-4B cross-encoder reranking.

Fig. 6 shows the retrieval system interface in operation. A natural-language query ("Child is jumping on the trampoline") retrieves and ranks semantically matched collages from across five sessions, displaying the full VLM-generated evidence text and reranker score alongside the 3×3 collage image.

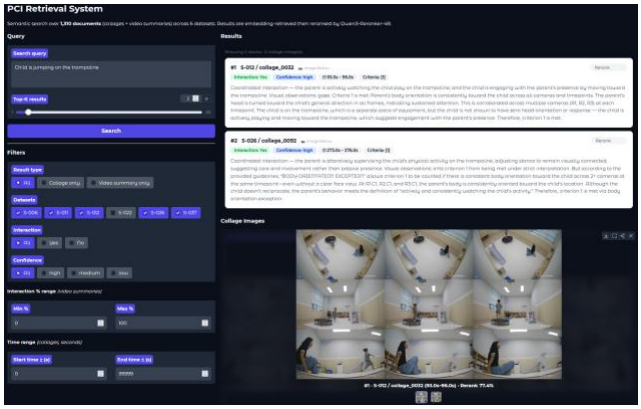



Figure 6. PCI Retrieval System interface. A natural-language query ("Child is jumping on the trampoline") retrieves semantically matched collages ranked by Qwen3-Reranker-4B. Each result card shows session ID, collage ID, interaction label, confidence, time window, criteria met, and the VLM-generated evidence text. The 3×3 spatiotemporal collage image is rendered below with its rerank score.

## V. Experiments and Results

### A. Experimental Setup

All VLM inference runs on a single NVIDIA GPU (24 GB VRAM) in bfloat16 precision. We use Qwen3-VL-8B-Instruct in training-free mode with the structured five-criterion prompt described in Section IV-C. No model weights are updated; the ten in-context calibration examples embedded in the prompt are expert-constructed and entirely independent of all evaluation sets. The same prompt is applied identically to all six sessions without session-specific adaptation.

### B. Evaluation Protocols

Primary metric: G-mean = √(TPR×TNR), which assigns 0 to any degenerate classifier (all-Yes or all-No). Secondary metrics: accuracy, precision, recall, and F1. All metrics are computed per session and then macro-averaged across the six sessions. We also report per session results for our best model.

### C. Main Classification Results

Table II reports macro-averaged metrics across all six sessions for each model. Qwen3-VL-8B-Instruct is the only model with meaningful discriminative ability: G-mean = 0.77 versus G-mean = 0.00 for Gemma-3-4B and Aya-Vision-8B, two models that collapse to always predicting 'Yes'. InternVL3-8B and Qwen3-VL-8B-Thinking show intermediate behavior; Nemotron-8B collapses to almost always predicting 'No'.

Table II. Overall macro-averaged metrics (n = 1,304 collages). G-mean is the primary metric. Bold = best. Gemma-3-4B and Aya-Vision-8B both achieve G-mean = 0.00.

| Model | Acc | Prec | Rec | F1 | G-mean |
|---|---|---|---|---|---|
| **Qwen3-VL-8B** | **79.75** | **94.40** | **80.89** | **87.12** | **0.771** |
| Qwen3-VL-8B-Think. | 83.64 | 87.97 | 93.47 | 90.64 | 0.522 |
| InternVL3-8B | 84.66 | 85.37 | 98.82 | 91.60 | 0.253 |
| Gemma-3-4B | 84.66 | 84.66 | 100.0 | 91.69 | 0.000 |
| Aya-Vision-8B | 84.66 | 84.66 | 100.0 | 91.69 | 0.000 |
| Nemotron-8B | 15.72 | 100.0 | 0.45 | 0.90 | 0.067 |

**Degenerate baselines.** Gemma-3-4B and Aya-Vision-8B predict 'Yes' for every collage. High nominal F1 scores (91.69%) disguise complete failure to detect non-interaction periods, a critical capability for clinical use.

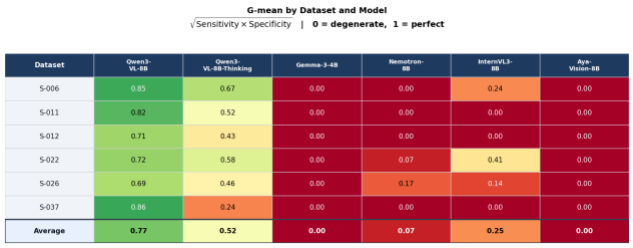



Figure 7. G-mean per model per session (green = high, red = low). Qwen3-VL-8B-Instruct consistently achieves the highest G-mean. Values of 0.00 indicate degenerate performance.

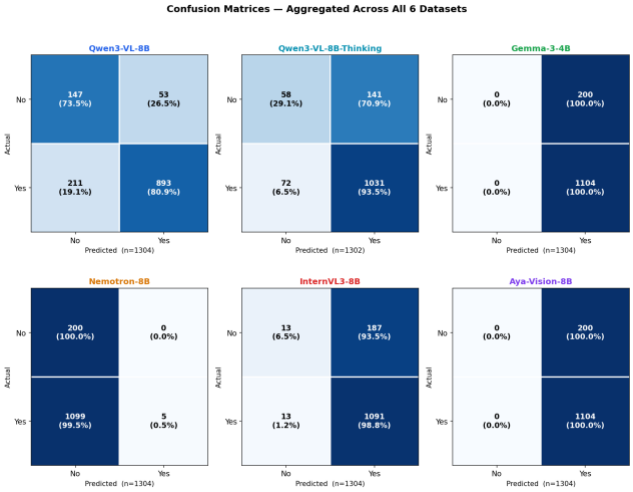



Figure 8. Per-model aggregated confusion matrices (row-normalized). Qwen3-VL-8B-Instruct is the only model that achieves non-trivial specificity.

### *D.* Per-Session Results

Table III provides per-session results for Qwen3-VL-8B-Instruct. Performance is highest on S-006 (G-mean = 0.853) and S-037 (G-mean = 0.859), both with the highest interaction prevalence. Performance is lowest on S-026 (G-mean = 0.686), the most balanced session (72.2% positive).

Table III. Qwen3-VL-8B-Instruct per-session metrics.

| Session | n | Acc | Prec | Rec | F1 | G-mean |
|---|---|---|---|---|---|---|
| S-006 | 189 | 78.84 | 99.25 | 77.33 | 86.93 | 0.853 |
| S-011 | 198 | 81.82 | 96.53 | 81.76 | 88.54 | 0.820 |
| S-012 | 197 | 81.73 | 94.90 | 84.18 | 89.22 | 0.711 |
| S-022 | 295 | 76.61 | 89.81 | 79.40 | 84.28 | 0.725 |
| S-026 | 194 | 66.49 | 86.41 | 63.57 | 73.25 | 0.686 |
| S-037 | 231 | 92.21 | 98.02 | 93.40 | 95.65 | 0.859 |
| **Average** | **-** | **79.75** | **94.40** | **80.89** | **87.12** | **0.771** |

### *E.* Instruct vs. Thinking Mode Ablation

Table IV compares Qwen3-VL-8B-Instruct against Qwen3-VL-8B-Thinking. The Thinking mode achieves higher recall (93.47% vs. 80.89%) but dramatically lower specificity, reducing G-mean from 0.771 to 0.522. This is most pronounced in S-037, where the Thinking model approaches a degenerate always-Yes behavior (G-mean = 0.236) despite the Instruct model achieving G-mean = 0.859.

Table IV. Instruct vs. Thinking mode comparison.

| Mode | Acc | Recall | Specificity | F1 | G-mean |
|---|---|---|---|---|---|
| **Instruct** | **79.75** | **80.89** | **~73%** | **87.12** | **0.771** |
| Thinking | 83.64 | 93.47 | ~29% | 90.64 | 0.522 |
| Delta | +3.89 | +12.58 | −44 | +3.52 | −0.249 |

The specificity degradation in Thinking mode is consistent with known behaviors of chain-of-thought reasoning in VLMs [38]: the model accumulates visual evidence incrementally and at each step finds some criterion that can be loosely interpreted as met. Our structured, evidence-quality-gated Instruct prompt effectively counteracts this accumulation bias.

### *F.* Overall Aggregated Performance

Fig. 9 shows the pooled confusion matrix for Qwen3-VL-8B-Instruct aggregated across all six sessions.

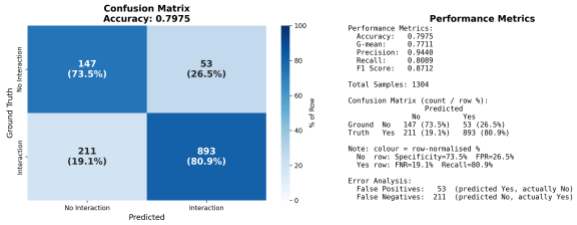



Figure 9. Aggregated confusion matrix for Qwen3-VL-8B-Instruct (n = 1,304). Each cell shows raw count and row-normalized percentage. The model achieves 80.9% recall and meaningful specificity compared to degenerate baselines.

### *G.* LLM-as-Judge Reasoning Quality

Figs. 10–12 show the LLM-as-Judge score distributions across the six sessions. Correct predictions (TP and TN) consistently receive scores of 6–9, while incorrect predictions (FP and FN) receive scores of 1–4, validating the judge's two-level rubric design. S-026 shows more variable TN scores, consistent with its more ambiguous non-interaction periods.

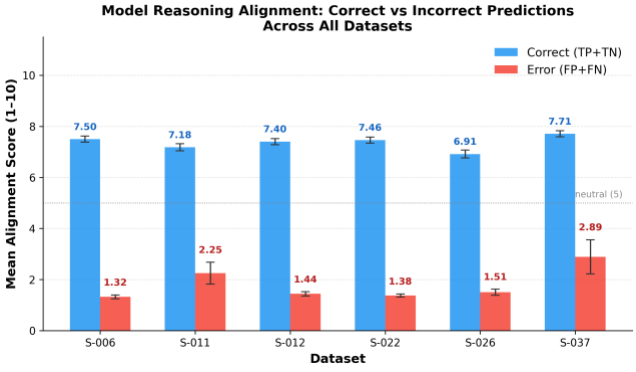



Figure 10. Correct vs. incorrect LLM-as-Judge scores per session (model reasoning mode). Correct predictions score 6–9; incorrect score 1–4.

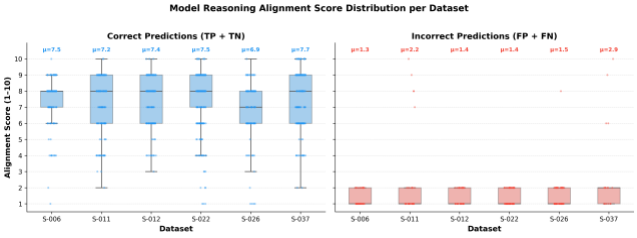



Figure 11. Judge score distribution by session. The bimodal pattern confirms the rubric's sensitivity to reasoning quality.

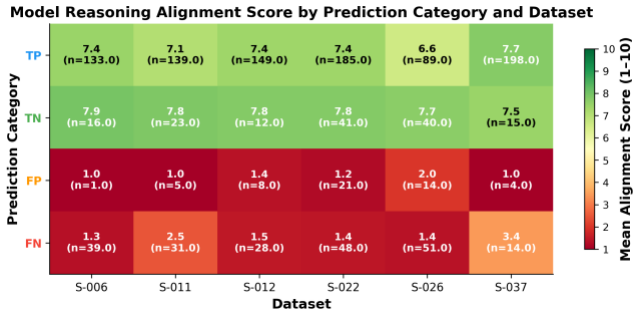



Figure 12. LLM-as-Judge score heatmap by prediction category (TP, TN, FP, FN). False positives and false negatives consistently receive low scores.

### *H.* Error Analysis

Qualitative analysis of FP (predicted Yes, GT=No) and FN (predicted No, GT=Yes) reveals three recurring patterns:

**Bin/basket gaze confusion (FP).** When both parent and child lean toward the same bin from opposite sides, the overhead camera creates a visual appearance of mutual gaze even though both are attending to the bin's contents. Our prompt includes an explicit exclusion rule, but overhead-angle geometry occasionally still triggers false criterion 3 claims.

**Corroboration gate misapplication (FN).** In sessions where one camera provides clear corroborated evidence across two timepoints, the model occasionally fails to recognize that same-camera consecutive-timepoint confirmation is sufficient, erroneously requiring multi-camera evidence.

**Physical handoff misclassification (FN).** Direct object handoffs from parent to child, which clearly satisfy Criterion 4 (physical contact via hand-to-hand object transfer), are occasionally classified as No when the handoff is brief and occurs in only one timepoint.

## VI. Discussion

### *A.* Effectiveness of Structured Prompting

Our results demonstrate that carefully crafted criterion-grounded prompting can achieve meaningful discriminative performance on a task traditionally requiring weeks of specialist training. The key design decisions are: (1) the spatiotemporal collage representation, which makes multi-view temporal information available in a single forward pass without requiring any explicit fusion mechanism; (2) the evidence quality gate, which counteracts the model's tendency to claim criteria based on ambiguous single-camera evidence; and (3) explicit exclusion rules covering known failure modes (bin confusion, parallel play, child on playground equipment).

### *B.* Why Standard Accuracy and F1 Are Insufficient

A critical finding is that standard accuracy (84.66%) and F1 (91.69%) are indistinguishable between our model and completely degenerate baselines on this dataset. This illustrates why imbalance-robust metrics are essential for behavioral observation tasks, where the clinically important capability is detecting non-interaction periods, the minority class. G-mean correctly assigns G-mean = 0.00 to degenerate classifiers while rewarding balanced performance.

### *C.* Instruct vs. Thinking: Prompt Engineering Implications

The contrast between Instruct and Thinking modes reveals a fundamental challenge in applying chain-of-thought reasoning to in-context prompted behavioral coding: the open-ended nature of CoT allows the model to accumulate and rationalize evidence toward a positive conclusion, systematically under-weighting corroboration requirements. This finding suggests that for behavioral observation tasks with strict evidence standards, concise instruction-following prompts with explicit gating rules outperform unconstrained reasoning chains.

### *D.* Limitations

Several limitations should be noted. First, ground truth was annotated by a single researcher without formal inter-rater reliability assessment; future work should use multi-rater agreement protocols (e.g., Cohen's κ or ICC). Second, prompt engineering was conducted iteratively on one session (S-037), which may introduce overfitting to that session's characteristics. Third, the CCTV setup uses fixed camera positions; the approach may not generalize to free-roaming footage without adaptation. Fourth, the pipeline processes pre-recorded footage at ≈40–60 collages/hour; real-time deployment would require optimization.

### *E.* Future Directions

Promising future directions include: (1) supervised fine-tuning of the VLM on labeled behavioral observation data using our training-free predictions as weak supervision; (2) extending the collage representation to longer windows (e.g., 9-second temporal grid) to capture slower behavioral dynamics; (3) integrating audio channels for verbal interaction detection (a criterion not yet modeled); and (4) adapting the retrieval system for comparative analysis across multiple research sites and dyads.

## VII. Conclusion

We presented a training-free automated parent-child interaction detection framework based on structured spatiotemporal collage prompting of Qwen3-VL-8B-Instruct, evaluated across 1,304 three-second collages from six naturalistic play sessions. Our model achieves a G-mean of 0.77, the only model with meaningful discriminative ability among six evaluated models, demonstrating that carefully engineered criterion-grounded VLM prompting can approach expert-level binary behavioral classification without any task-specific training.

Beyond the classification task, we demonstrated two complementary downstream modules: (1) an LLM-as-Judge pipeline providing automated reasoning quality assessment on a 1–10 factual consistency scale, correctly distinguishing correct from incorrect predictions; and (2) an Analysis Retrieval system enabling semantic search over the complete observation archive. Together, these form a complete, deployable pipeline from raw multi-camera video to a searchable, interpretable interaction archive.

Our findings highlight important methodological lessons: standard accuracy and F1 are insufficient for imbalanced behavioral coding tasks; G-mean should be the primary metric; and chain-of-thought reasoning modes can paradoxically degrade specificity in tasks with strict corroboration requirements. We release our prompt design and evaluation framework to support reproducible automated behavioral observation research.

## Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No RS-2023-00218176).

## References

[1] S. H. Landry, K. E. Smith, P. R. Swank, and C. L. Miller-Loncar, "Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills," Developmental Psychology, vol. 42, no. 4, pp. 627–642, 2006.

[2] B. Hart and T. R. Risley, Meaningful Differences in the Everyday Experience of Young American Children. Paul H. Brookes, 1995.

[3] M. Tomasello, Constructing a Language: A Usage-Based Theory of Language Acquisition. Harvard University Press, 2003.

[4] S. Eyberg and E. A. Robinson, "Parent-child interaction training: Effects on family functioning," J. Clinical Child Psychology, vol. 11, no. 2, pp. 130–137, 1982.

[5] S. M. Eyberg, M. M. Nelson, M. Duke, and S. R. Boggs, Manual for the Dyadic Parent-Child Interaction Coding System, 3rd ed. PCIT International, 2005.

[6] R. Bakeman and J. M. Gottman, Observing Interaction: An Introduction to Sequential Analysis. Cambridge University Press, 1986.

[7] T. Pfister, J. Charles, and A. Zisserman, "Flowing ConvNets for human pose estimation in videos," in Proc. ICCV, 2015.

[8] OpenAI, "GPT-4V(ision) technical report," arXiv:2303.08774, 2023.

[9] H. Liu, C. Li, Q. Wu, and Y. J. Lee, "Visual instruction tuning," in Proc. NeurIPS, 2023.

[10] J. Bai et al., "Qwen-VL: A versatile vision-language model," arXiv:2308.12966, 2023.

[11] B. Karaca, A. A. Salah, J. Denissen, R. Poppe, and S. M. C. de Zwarte, "Survey of automated methods for nonverbal behavior analysis in parent-child interactions," in Proc. IEEE FG 2024. DOI: 10.1109/FG59268.2024.10582009.

[12] M. Ahmadi, M. Topper, and M. Huber, "Assessment of parent–child interaction quality from dyadic dialogue," Applied Sciences, vol. 13, no. 20, p. 11129, 2023.

[13] H. Kaur, A. Gupta, and A. Vats, "Parent-child interaction assessment using computer vision: A systematic review," J. Medical Systems, vol. 44, no. 3, 2020.

[14] A. Jebeli, L. K. Chen, K. Guerrerio, S. Papparotto, L. Berlin, and B. J. Harden, "Quantifying the quality of parent-child interaction through machine-learning based audio and video analysis: Towards a vision of AI-assisted coaching support for social workers," ACM J. Computing and Sustainable Societies, vol. 2, no. 1, art. 6, 2024. DOI: 10.1145/3617693.

[15] T. Weng, L. Bravo-Sánchez, M. Rosen, and S. Yeung, "Artificial intelligence-powered 3D analysis of video-based caregiver-child interactions (HARMONI)," Science Advances, vol. 11, no. 8, eadp4422, 2025.

[16] P. Wang et al., "Qwen2-VL: Enhancing vision-language model's perception of the world at any resolution," arXiv:2409.12191, 2024.

[17] Qwen Team, "Qwen3-VL technical report," arXiv:2511.21631, 2025.

[18] Z. Chen et al., "InternVL: Scaling up vision foundation models," in Proc. CVPR, 2024.

[19] M. Maaz, H. Rasheed, S. Khan, and F. S. Khan, "Video-ChatGPT: Towards detailed video understanding," in Proc. ACL 2024.

[20] Z. Cheng et al., "VideoLLaMA 2: Advancing spatial-temporal modeling," arXiv:2406.07476, 2024.

[21] K. Li et al., "VideoChat: Chat-centric video understanding," arXiv:2305.06355, 2023.

[22] S. Ren et al., "TimeChat: A time-sensitive multimodal LLM for long video understanding," in Proc. CVPR, 2024.

[23] C. Fu et al., "Video-MME: The first-ever comprehensive evaluation benchmark," arXiv:2405.21075, 2024.

[24] W. Shi, V. H. Le, and K. T. W. Choo, "Towards multimodal large-language models for parent-child interaction: A focus on joint attention," in CHI EA 2025. arXiv:2502.19877

[25] W. Shi and K. T. W. Choo, "Towards aligning multimodal LLMs with human experts: A focus on parent-child interaction," arXiv:2511.04366, 2025. Accepted at CHI 2026.

[26] W. Shi and K. T. W. Choo, "Human-AI alignment of multimodal large language models with speech-language pathologists in parent-child interactions," arXiv:2506.05879, 2025..

[27] CollagePrompt Authors, "CollagePrompt: A benchmark for budget-friendly visual recognition with GPT-4V," arXiv:2403.11468, 2024.

[28] P. Li, H. Lu, R. Poppe, and A. A. Salah, "Automated detection of joint attention and mutual gaze in free play parent-child interactions," in Proc. ICMI 2023. DOI: 10.1145/3610661.3616234.

[29] C. Beyan, A. Vinciarelli, and A. Del Bue, "Co-located human–human interaction analysis using nonverbal cues: A survey," ACM Computing Surveys, vol. 56, no. 7, 2024.

[30] L. Zheng et al., "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena," in Proc. NeurIPS. arXiv:2306.05685, 2023.

[31] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Proc. NeurIPS, vol. 33, 2020.

[32] Qwen Team, "Qwen3 embedding: Advancing text embedding and reranking," arXiv:2506.05176, 2025.

[33] Z. Zhou et al., "LLM×MapReduce: Simplified long-sequence processing using large language models," arXiv:2410.09342, 2024.

[34] M. Kubat and S. Matwin, "Addressing the curse of imbalancedness: One-sided selection," in Proc. ICML, pp. 179–186, 1997.

[35] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics," Pattern Recognition, vol. 91, pp. 216–231, 2019.

[36] Qwen Team, "Qwen3 technical report," arXiv:2505.09388, 2025.

[37] Gemma Team, "Gemma: Open models based on Gemini research and technology," arXiv:2403.08295, 2024.

[38] J. Wei et al., "Chain-of-thought prompting elicits reasoning in large language models," in Proc. NeurIPS, 2022.

# Automated Scoring of Coping on Draw-a-Person-in-the-Rain (DAPR) Sketches Using YOLO Segmentation and Multi-Task Learning

*L. Senghak*[1], *Syed Ali Naqi Raza*[2], *T. Sokea*[2], *P. Samnang*[1], *Chomyong Kim*[2], and *Yunyoung Nam*[3]

*1 Department of Computer Science and Engineering, Soonchunhyang University, Asan 31538, Republic of Korea*

*2 ICT Convergence Research Center, Soonchunhyang University, Asan 31538, Republic of Korea*

*3 Emotional and Intelligent Child Care Convergence Center, Soonchunhyang University, Asan 31538, Republic of Korea*

**Corresponding Author:** Yunyoung Nam. Email: ynam@sch.ac.kr

**Abstract** — The "Draw a Person in the Rain" (DAPR) test is a psychological assessment tool evaluating an individual's vulnerability to environmental stress and the effectiveness of their coping mechanisms. Traditionally, manual scoring of these sketches is time-consuming and subjective. To address these clinical challenges, this paper presents an automated scoring system for the DAPR test using a hybrid computer vision pipeline. The system uses YOLOv11-seg combined with OpenCV heuristics to extract structural features and Region of Interest (ROI) crops from sketches. Building upon this foundation, we implement an Improved ROI-CNN architecture that processes three ROI crops (person, umbrella, face) through a shared from-scratch convolutional backbone. The model employs criteria-specific routing, where each scoring head receives only the image features and structured heuristics logically relevant to its criterion, encoding domain knowledge directly into the architecture. Per-ROI projections rebalance image features against sparse structured inputs, and homoscedastic uncertainty weighting (Kendall et al., 2018) automatically balances losses across eight criteria of varying difficulty. Using 5-fold cross-validation on 169 ground-truth annotated sketches, the Improved ROI-CNN achieves 81.7% overall exact-match accuracy across all criteria, a substantial improvement over the 69.4% achieved by the rule-based heuristic baseline.

**Keywords** — DAPR; YOLO11; Image Segmentation; Automated Scoring; Psychological Assessment; ROI-CNN; Multi-Task Learning; Ordinal Regression; Computer Vision.

# I. Introduction

## Background of the selected topic:

### *Why does this problem need solving?*

Historically, psychologists and trained clinicians have had to review and score these drawings using standardized rubrics manually. As highlighted in recent literature [3], this manual process is highly subjective, prone to inter-rater reliability issues, and difficult to scale across large populations such as educational institutions. It demands considerable time and resources, often requiring costly and labor-intensive inter-rater reliability assessments among multiple experts to validate the scores.

### *What is the importance of this topic?*

Automating this scoring process provides a consistent, objective, and rapid assessment tool. It can act as a reliable preliminary screening mechanism in clinical psychology, assisting practitioners by standardizing evaluations and reducing human fatigue. Previous automated DAPR systems, such as AlphaDAPR, have successfully demonstrated this potential by utilizing YOLO-v5 [2] for baseline object detection. Our research advances this foundation by implementing the newer, more precise YOLO11-seg architecture. By transitioning from simple bounding box detection to pixel-level mask generation, this approach achieves the superior segmentation required to accurately parse complex, highly variable hand-drawn sketches.

## Objectives of your work

- To develop a robust object detection and segmentation pipeline capable of parsing complex, highly variable hand-drawn sketches.
- To automate the extraction of explicit scoring criteria (CR1–CR11) outlined in the DAPR Protection & Coping Scoring manual.
- To design and implement an Improved ROI-CNN architecture utilizing criteria-specific routing, per-ROI feature projections, and homoscedastic uncertainty weighting for multi-task learning.

# II. Datasets

## Description of the dataset used:

***General description of the data:***

The dataset consists of hand-drawn sketches of the DAPR test. The images vary widely in stroke darkness (from faint pencil to dark ink) and background quality. A custom data preparation pipeline was built to standardize annotations, extract masks and bounding boxes, and normalize class labels. To support the complete scoring rubric, the dataset's ontology includes 7 primary classes for YOLO detection: person, umbrella, clouds, face, indirect_protection, puddle, and rain.

***How is the data obtained/constructed?***

Data was collected from physical sketch submissions provided by subjects. These physical copies were scanned or photographed, digitized, and then manually annotated using bounding boxes and segmentation masks. The corresponding ground-truth scores for the 11 clinical criteria were assigned by trained raters.

***Data partitioning:***

The dataset contains 196 images in total. Images with IDs 1–169 carry human ground-truth (GT) labels for all scoring criteria. Images with IDs 170–196 carry OpenCV-predicted pseudo-labels. For deep learning evaluation, rigorous 5-fold stratified cross-validation is applied on the 169 GT images only. In each fold, the training set consists of approximately 135 GT samples plus all 27 pseudo-label samples, while the validation set consists of approximately 34 GT samples. Pseudo-label samples are never included in validation, ensuring evaluation integrity.

# III. Proposed Methods

Table 1 presents the 11 coping scoring criteria from the DAPR Protection & Coping manual. The deep learning model addresses 8 of these criteria (CR1–CR4, CR6, CR8, CR9, CR11). CR5 (indirect protection adequacy) is not yet automated, and CR10 (completeness) is scored via a separate heuristic. **CR7 (facial expression) is handled by a dedicated EfficientNet-B0 classifier pretrained on ImageNet and fine-tuned on 450 face crops (360 training, 90 validation) for 3-class mood classification (poor/neutral/good), achieving 0.956 validation accuracy.** This classifier operates independently from the main ROI-CNN, taking YOLO-detected face crops as input.

| No. | Item | Scoring Criteria |
|---|---|---|
| 1 | Number of direct protective objects | 1: None 2: 1 item 3: ≥ 2 items |
| 2 | Number of indirect protective objects | 1: None 2: 1 item 3: ≥ 2 items |
| 3 | Total number of protective objects | 1: None 2: 1–2 items 3: ≥ 3 items |
| 4 | Adequacy of direct protection | 1: None 2: Insufficient 3: Adequate |
| 5 | Adequacy of indirect protection | 1: None 2: Insufficient 3: Adequate |
| 6 | Facial coverage | 1: Completely covered 2: Partially covered 3: Fully visible |
| 7 | Facial expression | 1: Poor mood 2: Neutral 3: Good mood |

| | | |
|---|---|---|
| 8 | Size of the person | 1: Very small (≤2 in) 2: Average 3: Very large (≥6 in) |
| 9 | Position of the person | 1: Edge 2: Slightly off-center 3: Centered |
| 10 | Completeness of the person | 1: Upper body only 2: Lower body incomplete 3: Whole body |
| 11 | Stroke quality (line quality) | 1: Very shaky 2: Partially unstable 3: Smooth |

## Description of the proposed method(s)

The proposed system comprises two stages: a rule-based heuristic pipeline for structural feature extraction, followed by an Improved ROI-CNN that fuses visual features with these structured heuristics via multi-task learning.

***Stage 1: Heuristic Pipeline***

The existing system relies on YOLO11-seg for detection and mask generation, paired with specialized OpenCV pipelines for each criterion:

- **Facial Coverage (CR6):** Uses segmentation masks and pixel balance analysis to determine if a face is fully visible or obscured.
- **Position of Person (CR9):** Calculates the center of mass from the person segmentation mask to identify edge/center placements.
- **Size of the Person (CR8):** Converts pixel height to real-world inches, assuming a standard A4 paper scale.
- **Stroke Quality (CR11):** Analyzes contour angle variance and simplification ratios on the person's outline to detect shakiness.
- **Direct Protection (CR4):** Evaluates mask overlap ratios and vertical coverage between protective items and the subject.
- **Object Counts (CR1, CR2, CR3):** Counts YOLO detections for direct and indirect protection objects.

These heuristic scripts also generate 16 structured numeric features (e.g., overlap_ratio, person_height_inches, angle_variance, pixel_balance_ratio) that are passed to the deep learning model as auxiliary inputs.

***Stage 2: ROI-CNN Architecture***

To improve upon the heuristic baseline, we implemented an Improved ROI-CNN that processes three YOLO-cropped ROI images (person, umbrella, face) through a shared convolutional backbone, projects them into balanced feature subspaces, and routes them to criteria-specific fusion heads alongside selected structured features. The architecture is described in detail below.

## Detailed architecture and setting configurations

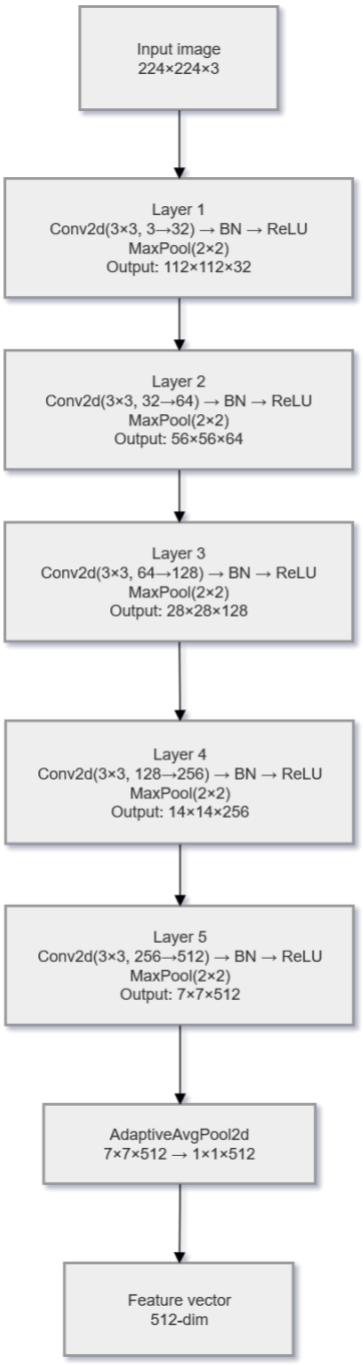
The Improved ROI-CNN architecture consists of five stages: a shared CNN backbone, per-ROI projections, criteria-specific routing, per-criteria fusion MLPs, and ordinal output heads. Learnable per-task loss weights automatically balance the multi-task training objective.

***SharedCNNBackbone (SimpleCNNBackbone):***

A custom 5-layer convolutional network processes each ROI crop (224×224×3).

Each layer follows the pattern Conv2d(3×3) → BatchNorm → ReLU → MaxPool(2×2), with channel progression 3 → 32 → 64 → 128 → 256 → 512.

An AdaptiveAvgPool2d reduces the final feature maps to a 512-dimensional vector per crop. The same backbone weights are shared across all three ROI types; at each forward pass, the three crops are stacked along the batch dimension and processed in a single call for efficiency.




### *Per-ROI Projections (512 → 128):*

Each ROI type (person, umbrella, face) has its own dedicated linear projection: Linear(512 → 128) → LayerNorm → ReLU → Dropout(0.1). Separate projections allow each ROI type to emphasize different aspects of the shared backbone features. Critically, this projection also rebalances the dimensionality: before projection, a criterion like CR1 received 1024 image dimensions alongside just 1 structured feature (umbrella_count), making the structured input nearly invisible. After projection, the ratio becomes 256 image dimensions to 1 structured feature, substantially increasing the relative signal strength of domain-engineered heuristics.

### *Criteria-Specific Routing:*

Rather than feeding all features to every scoring head, each criterion receives only the ROI features and structured inputs that are logically relevant to its scoring rubric. This encodes domain knowledge directly into the architecture, which is essential given the small training set (~108 samples per fold). The routing configuration is as follows:

| Criterion | Scoring Target | ROI Inputs | Struct. Feats | Input Dim |
|---|---|---|---|---|
| CR1 | Direct protection count | Person + Umbrella | 1 | 257 |
| CR2 | Indirect protection count | Person + Umbrella | 1 | 257 |
| CR3 | Total protection count | Person + Umbrella | 1 | 257 |
| CR4 | Adequacy of protection | Person + Umbrella | 3 | 259 |
| CR6 | Facial coverage | Face + Umbrella | 2 | 258 |
| CR8 | Person size | Person only | 1 | 129 |
| CR9 | Position of a person | Person only | 2 | 130 |
| CR11 | Stroke quality | Person + Face | 5 | 261 |

### *Per-Criteria Fusion MLP:*

Each criterion has its own single-layer fusion network with independent weights: Linear(input_dim → 64) → LayerNorm → ReLU → Dropout(0.5). A single hidden layer was chosen over two to reduce overfitting on the small training set. LayerNorm replaces BatchNorm because batch statistics are unreliable at the effective batch size of 16.

***Ordinal Output Heads:***

Each criterion outputs 2 logits representing the ordinal thresholds P(score ≥ 2) and P(score ≥ 3). The final score is decoded as: score = 1 + I($\sigma(t_1)$ > 0.5) + I($\sigma(t_2)$ > 0.5), where σ is the sigmoid function. This ordinal encoding preserves the natural ordering of scores, unlike standard multi-class classification, which treats all misclassifications equally. The loss function is Binary Cross-Entropy with logits (BCEWithLogitsLoss).

***Uncertainty-Weighted Multi-Task Loss:***

Following Kendall et al. [6], each criterion is assigned a learnable log-variance parameter log_var_t, initialized to zero (equal weights). The total loss is:

$$L = \Sigma_t \left[ \exp(-\log_var_t) \times L_t + 0.5 \times \log_var_t \right]$$

The first term weights each task's loss by its learned precision ($1/\sigma^2$). The second term regularizes the log-variance, preventing the trivial solution of pushing all weights to zero. During training, the model automatically learns to down-weight criteria with high irreducible noise (e.g., subjectively-rated criteria with low inter-rater agreement) and up-weight criteria with clean, learnable signals.

***Training Configuration:***

| Parameter | Value | Description |
|---|---|---|
| Epochs | 200 (max) | With early stopping |
| Learning Rate | 1e-4 | Adam optimizer |
| Scheduler | Cosine Annealing | Decays LR over epochs |
| Weight Decay | 1e-4 | L2 regularization |
| Early Stopping | 15 checks (75 epochs) | Evaluated every 5 epochs |
| Gradient Clipping | max_norm = 1.0 | Prevents gradient explosion |
| Batch Size | 16 | Per training step |
| ROI Projection Dim | 128 | Per-ROI feature size |
| Fusion Dropout | 0.5 | In fusion MLP |

## Reasons behind choosing/designing the method(s)

- **From-scratch CNN over pretrained backbone:** ImageNet-pretrained models (e.g., EfficientNet-B0) were tested but did not outperform the from-scratch backbone. This is attributed to the domain gap: ImageNet contains photographs of real-world objects with rich color and texture, whereas DAPR sketches are sparse pencil drawings with no color information. The visual statistics are fundamentally different, making pretrained low-level filters (designed for photographic textures) poorly suited for sketch analysis.
- **Criteria-specific routing:** With only ~108 training samples per fold, the model cannot afford to discover which features are relevant to each criterion on its own. Routing encodes domain knowledge — for example, CR6 (facial coverage) needs the face and umbrella crops but not the person crop, while CR8 (person size) needs only the person crop. This dramatically reduces the search space for each scoring head.

- **Per-ROI projections:** Before projection, structured features were numerically invisible (e.g., 1 feature out of 1025 input dimensions for CR1). Projecting each ROI from 512 to 128 dimensions rebalances the input so that domain-engineered heuristics carry meaningful weight in the fusion.
- **Uncertainty weighting over equal loss weights:** The 8 criteria have different levels of label noise and difficulty. Equal weighting forces the shared backbone to compromise equally across all tasks. Uncertainty weighting [6] automatically allocates gradient budget to tasks where loss reduction reflects genuine learning progress, rather than memorization of noisy labels.
- **Ordinal regression:** Standard 3-class softmax treats "predicting 1 when truth is 3" identically to "predicting 2 when truth is 3." Ordinal encoding via cumulative thresholds preserves the natural score ordering, which is clinically meaningful in psychological assessment.
- **Otsu's thresholding:** Chosen over fixed thresholds to reliably handle the massive variance in sketch line darkness across different drawing mediums [4].
- **Center of mass over bounding box:** Provides better accuracy for position scoring on sketches with dynamic poses (e.g., walking, reaching) compared to simple bounding box centers.

# IV. Experiments and Results

## Experimental Setup

### *Heuristic Pipeline:*

The YOLO11-seg object detection model was configured with 7 classes (person, umbrella, clouds, face, indirect_protection, puddle, rain). Python pipelines applied morphological operations and contour analysis over YOLO mask outputs. For the rule-based heuristic evaluations, 830 valid instances were used to evaluate pure algorithmic accuracy across the 5 automated criteria (CR4, CR6, CR8, CR9, CR11).

### *Improved ROI-CNN Training:*

The Improved ROI-CNN was trained using single-stage end-to-end optimization. The from-scratch SimpleCNNBackbone, per-ROI projections, fusion MLPs, ordinal output heads, and learnable log-variance parameters were all trained jointly from initialization using the Adam optimizer with a learning rate of 1e-4 and cosine annealing. Training ran for up to 200 epochs per fold with early stopping (patience of 15 evaluation checks at 5-epoch intervals). Gradient clipping (max_norm = 1.0) was applied to prevent instability.

### *Data Partitioning:*

5-fold stratified cross-validation was applied on the 169 human-annotated ground-truth images, stratified by CR1 scores. In each fold, the training set comprised approximately 135 GT images plus 27 pseudo-labeled images (scored by OpenCV heuristics). The validation set comprised approximately 34 GT-only images. Pseudo-labeled samples were restricted to the training set to preserve evaluation integrity.

## Evaluation Protocols

- **Exact Match Accuracy:** The percentage of predictions that perfectly match the human-annotated ground truth (1, 2, or 3).
- **±1 Tolerance Accuracy:** The percentage of predictions within one point of the ground truth. In ordinal psychological scoring, adjacent errors are clinically preferred over distant errors.

## Results

### *OpenCV Heuristic Baseline (5 criteria, 830 instances):*

| Criteria | Exact Match | ±1 Tolerance | MAE |
|---|---|---|---|
| CR1 Direct Count | 129/169 (76.3%) | 160/169 (94.7%) | **0.290** |

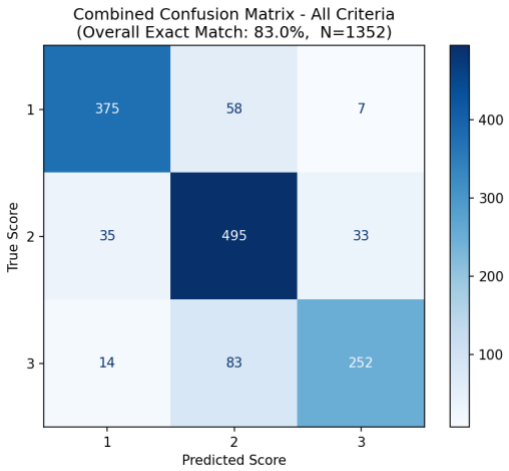
| | | | |
|---|---|---|---|
| CR2 Indirect Count | 157/169 (92.9%) | 168/169 (99.4%) | **0.077** |
| CR3 Total Protection | 136/169 (80.5%) | 165/169 (97.6%) | **0.219** |
| CR4 Direct Protection | 120/169 (71.0%) | 152/169 (89.9%) | **0.391** |
| CR6 Facial Coverage | 114/169 (67.5%) | 166/169 (98.2%) | **0.343** |
| CR7 Facial Expression | 171/196 (87.2%) | 184/196 (93.9%) | **0.189** |
| CR8 Size of Person | 128/169 (75.7%) | 169/169 (100.0%) | **0.243** |
| CR9 Position | 113/169 (66.9%) | 167/169 (98.8%) | **0.343** |
| CR11 Stroke Quality | 114/169 (67.5%) | 169/169 (100.0%) | **0.325** |
| **Total** | **1182/1548 (76.4%)** | **1500/1548 (96.9%)** | **0.267** |

### *Improved ROI-CNN Performance Metrics (8 criteria, 5-fold CV on 169 GT images):*

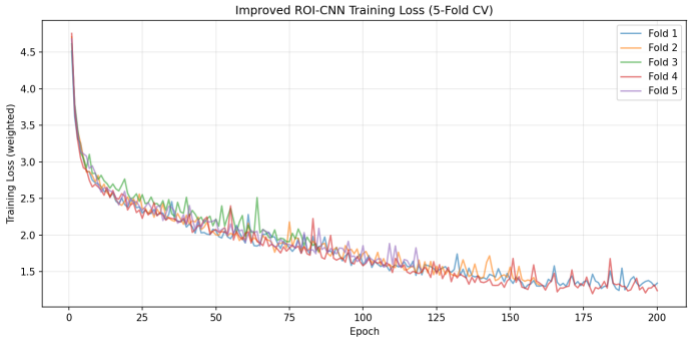
| Criteria | Exact Match | ±1 Tolerance | Precision | Recall | F1 |
|---|---|---|---|---|---|
| CR1 Direct Count | 79.9% | 97.0% | 69.5% | 79.9% | 73.9% |
| CR2 Indirect Count | 94.7% | 100.0% | 90.7% | 94.7% | 92.7% |
| CR3 Total Protection | 82.2% | 98.2% | 76.9% | 82.2% | 79.5% |
| CR4 Direct Protection | 83.4% | 93.5% | 79.0% | 83.4% | 81.1% |
| CR6 Facial Coverage | 88.8% | 100.0% | 85.1% | 88.8% | 85.8% |
| CR8 Size of Person | 88.8% | 100.0% | 89.7% | 88.8% | 88.4% |
| CR9 Position | 75.1% | 99.4% | 71.2% | 75.1% | 71.4% |
| CR11 Stroke Quality | 71.0% | 99.4% | 69.2% | 71.0% | 64.6% |
| **Total** | **82.9%** | - | - | - | - |

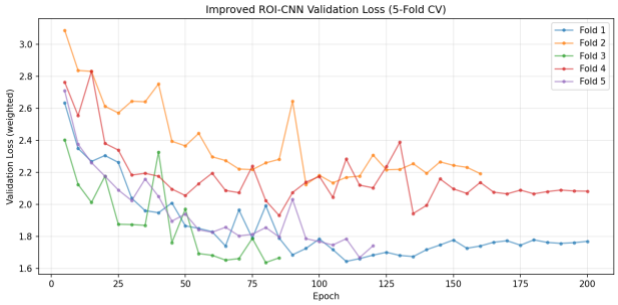
*Note: The Learned Weight column can be populated with the actual uncertainty weights once training is finalized. A higher weight indicates the model finds the task's loss signal more reliable.*

### *Confusion Matrix:*




### *Train Loss and Validation Loss:*






## Discussion

The Improved ROI-CNN achieves 82.9% overall exact-match accuracy across all 8 criteria, representing a substantial improvement over the 76.4% overall heuristic baseline. The ±1 tolerance accuracy of 97.8% demonstrates that misclassifications are predominantly adjacent (e.g., predicting score 2 when the truth is 3).

For criteria shared between the heuristic and CNN pipelines (CR4, CR6, CR8, CR9, CR11), the deep learning model achieved significant gains. The largest improvements were seen on CR6 (Facial Coverage) with an exact match of 88.8% compared to the heuristic 67.5% (+21.3%) and CR8 (Size of Person), improving from 75.7% to 88.8% (+13.1%). These gains in criteria involving subjective judgments about spatial adequacy (CR6) and contextual cues (CR8) suggest the CNN captures nuances that rule-based thresholds handle poorly. CR4 (Direct Protection) also showed strong improvement, achieving 83.4% over the heuristic 71.0%.

CR9 (Position, 75.1%) and CR11 (Stroke Quality, 71.0%) remain the most challenging criteria. CR9's difficulty is partially attributable to the horizontal flip augmentation used during training, which can corrupt position-dependent labels. CR11's low accuracy likely reflects inherent label noise: stroke quality is the most subjective criterion, and inter-rater agreement is expected to be lower than for objective criteria like object counting.

Analysis of the combined confusion matrix reveals a systematic central tendency bias: score 3 is under-predicted, with 27% of true score-3 cases misclassified as score 2. This is characteristic of class imbalance in the training data, where score 2 dominates. Future work should explore class-weighted loss functions to address this pattern.

The training loss curves show smooth convergence across all folds, with validation loss plateauing around epoch 50–75. The gap between training and validation loss after this point indicates mild overfitting, which is mitigated by the early stopping mechanism.

# V. Conclusion

## Summary

This paper presented an automated scoring system for the DAPR psychological assessment, combining YOLO11-seg instance segmentation with an Improved ROI-CNN for multi-task ordinal prediction. The system automates 8 of the 11 Protection & Coping scoring criteria. The heuristic pipeline achieved 76.4% exact-match accuracy using rule-based spatial analysis, while the Improved ROI-CNN raised this to 82.9% by fusing visual features from three ROI crops with structured heuristic outputs through criteria-specific routing and uncertainty-weighted multi-task learning.

## Final thoughts

The key architectural insight is that criteria-specific routing — encoding which features are relevant to which scoring criterion — is more effective than letting the model discover these associations from limited data. Combined with per-ROI projections for dimensional rebalancing and uncertainty weighting for automatic loss balancing, these design choices collectively address the fundamental challenge of training a multi-task model on fewer than 200 labeled samples. The near-perfect ±1 tolerance accuracy (97.8%) achieved with the final model, which yields 82.9% exact match accuracy, demonstrates that the system is viable as a preliminary clinical screening tool.

## Perspective/future work

- Address the central tendency bias (score-3 underestimation) through class-weighted BCE loss or oversampling strategies.

- Measure inter-rater agreement (Cohen's κ) per criterion to distinguish model limitations from label noise ceilings, particularly for CR9 and CR11.
- Automate CR5 (adequacy of indirect protection), the only remaining unscored criterion.
- Investigate Vision-Language Models (VLMs) as a complementary scoring modality for subjective criteria.
- Implement advanced instance segmentation models to accurately extract the full human silhouette for CR10 (completeness), enabling the analysis of complex body postures and precise structural boundaries.
- Automate CR5 (adequacy of indirect protection) by expanding the dataset to include diverse environmental shelters (e.g., building overhangs, canopies), facilitating the complex spatial reasoning required to determine full protection.

## Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176).

## References

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.

[2] J. Kim, J. Kang, T. Kim, H. Song, and J. Han, "AlphaDAPR: An AI-based Explainable Expert Support System for Art Therapy," in 28th International Conference on Intelligent User Interfaces (IUI '23), Sydney, NSW, Australia, March 2023.

[3] M. Yang, C. Park, T. Kim, H. Song, and J. Han, "CheckDAPR: An MLLM-based Sketch Analysis System for Draw-A-Person-in-the-Rain Assessments," in Proc. ACM, 2025, doi: 10.1145/3746252.3761571.

[4] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Trans. Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[5] J. Kim, M. Jo, S. Nam, Y. Lee, and Y. Baik, "Decoding stress and coping ability levels through drawing features in the Person-In-The-Rain test: An exploratory study using a digital drawing device," Computers in Human Behavior Reports, vol. 21, p. 100902, 2026.

[6] A. Kendall, Y. Gal, and R. Cipolla, "Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7482–7491.

[7] R. Caruana, "Multitask Learning," in Learning to Learn, Springer, 1998, pp. 95–133.

# Relationship between Sensory Processing Patterns, Autonomic Nervous System, and Occupations

Danbi Lee[1], Deukgeun Yoon[2], Dawon Jung[2], Hanju Kang[2], Minkyu Kim[2], Dabin Kim[2], Yoonseo Kim[2], Eun Young Kim[3, *]

[1, 2]*Department of Occupational Therapy, Soonchunhyang University, Asan, Republic of Korea*
[3] *Professor, Department of Occupational Therapy, Soonchunhyang University, Asan, Republic of Korea, Corresponding author*
**Contact: creativelee@sch.ac.kr, +82-41-530-1114*

*Abstract* **Purpose** This study aimed to examine the relationships among sensory processing patterns, autonomic nervous system, and occupations in young adults. **Methods** Fifty-one adults aged 18–25 years participated in this study. Heart rate variability (HRV), indicated by the Standard deviation of normal-to-normal intervals (SDNN), was obtained from resting electrocardiography (ECG). Interoception was assessed using the Heartbeat Tracking Task. Sensory processing patterns were evaluated using the Adolescent/Adult Sensory Profile (AASP), and occupational competence and value were measured using the Occupational Self Assessment (OSA). The collected data were analyzed using Pearson correlation analysis and multiple regression analysis with SPSS. **Results** The results indicated that among the sensory processing patterns, low registration was a key variable significantly associated with autonomic nervous system indicators and occupations. Higher low registration scores were associated with lower HRV ($r = -.317^{*}$) and higher interoceptive error scores ($r = .343^{*}$). In addition, multiple regression analysis revealed that the low registration pattern significantly and negatively predicted occupational competence ($\beta = -.448^{*}$). **Conclusion** This study empirically examined the relationships among sensory processing patterns, autonomic nervous system regulation, and occupations in young adults. In particular, low registration was significantly associated with autonomic nervous system indicators and occupational competence. These findings suggest that sensory processing characteristics in young adulthood may be closely related to physiological regulation and occupations in everyday life. Understanding individuals' sensory processing characteristics may therefore provide important foundational information for understanding their neural regulation and experiences of occupational performance.

## I. Introduction

Sensory processing refers to the process of perceiving, organizing, and interpreting sensory information to generate adaptive responses (Dunn et al., 2002; Kilroy et al., 2019), and individual differences exist in how these stimuli are noticed and modulated,
forming distinct sensory processing patterns (Harrold et al., 2024). Individuals differ in their neurological thresholds for detecting and responding to sensory stimuli in everyday life (Brown et al., 2001; Dunn, 2007). This range of responding is dependent on the functioning of the nervous system, and at the extremes of the neurological threshold continuum are habituation and sensitization, which are associated with the maturation of the central nervous system (Eeles et al., 2012).

The nervous system processes information from both external and internal sensory input from environments and maintains homeostasis (Porges, 2007). In particular, the autonomic nervous system regulates physiological responses through the interaction of the sympathetic and parasympathetic systems (Bayram-Weston et al., 2022). Autonomic nervous system function can be reflected by heart rate variability (HRV), which represents regulatory capacity and adaptability to environmental changes (Shaffer & Ginsberg, 2017; Grippo & Johnson, 2002). Previous studies have also reported that HRV, which reflects the efferent component of the ANS, is associated with cardiac interoceptive accuracy, which is related to the afferent component of the ANS, suggesting a link between autonomic regulation and the perception of internal bodily signals (Lischke et al., 2021).

Interoception refers to the process of sensing and interpreting internal bodily states and is closely related to emotion and self-regulation (Quigley et al., 2021; Tsakiris & Critchley, 2016). It contributes to homeostasis through regulation of the autonomic nervous system (Khalsa et al., 2018). ANS-related physiological responses may vary depending on sensory processing patterns (Brown et al., 2001). However, previous studies have primarily examined sensory processing, autonomic function, and interoception separately, and studies integrating these factors remain limited. In particular, there is a lack of research investigating how these physiological and sensory mechanisms are related to occupations.

Therefore, this study aimed to examine the relationships among sensory processing patterns, ANS-related physiological regulation, and interoception in young adults, and to investigate how sensory processing patterns predict perceived occupational experience.

## II. Material and Methods

### *A. Participants and Procedure*

Participants were undergraduate students aged 18–25 years recruited through the community of S University. Individuals who provided informed consent were included in the study. From an initial sample of 51 participants, 8 outliers were excluded. Accordingly, Pearson's correlation analysis was conducted for 43 participants regarding the AASP and SDNN. In addition, 9 participants were excluded from the Heartbeat Tracking Task analysis, resulting in a final sample of 34 participants for this analysis. Multiple regression analysis was performed on a final sample of 43 participants. A heartbeat perception task and Electrocardiography (ECG) recording were conducted in a sound-attenuated laboratory, followed by self-report questionnaire assessments. Data collection was conducted between September 1 and October 2, 2025.

### *B. ECG*

Heart rate was measured using Lead II of the bipolar standard limb lead configuration, with electrodes placed on the right wrist (–), left ankle (+), and left wrist (reference). Electrocardiography signals were recorded using the AIM Gen 2 Physiological Monitor (CGX Systems, USA) in a sound-attenuated environment.

#### *a. Heartbeat Tracking Task*

The Heartbeat Tracking Task was administered using the method developed by Schandry (1981). After electrode placement, participants sat comfortably and were given sufficient rest before measurement. During the task, physical movements or manipulations that could facilitate heartbeat perception, such as palpating the pulse or regulating breathing, as well as communication, were not allowed. Participants counted their heartbeats during three periods of 25, 35, and 45 seconds and verbally reported the counts immediately after each period. Auditory signals indicated the start and end of each period, and no information regarding period duration or task accuracy was provided. The reported heartbeat counts were compared with actual counts derived from electrocardiography signals to calculate the Heartbeat Error Score. For each period, the absolute difference between actual (HB_Actual) and reported (HB_Reported) heartbeats was divided by HB_Actual, and the values were summed across periods. Higher scores indicated greater error in heartbeat perception.

#### *b. Heart rate variability*

Heart rate variability (HRV) was derived from electrocardiogram (ECG) signals. Measurements were conducted in a stable seated position while participants remained at rest. ECG was recorded for 5 minutes, as recommended for short-term HRV analysis (Task Force of the European Society of Cardiology & the North American Society of Pacing and Electrophysiology, 1996). During recording, physical movement and communication were restricted to minimize potential interference with ECG signals. The HRV index SDNN was obtained using the analysis software integrated with the device.

### *C. AASP*

The Adolescent/Adult Sensory Profile (AASP) is a self-report measure developed based on Dunn's model of sensory processing to assess sensory processing patterns in daily life (Brown & Dunn, 2002). These patterns are categorized into four quadrants according to the interaction between neurological thresholds (high/low) and self-regulation strategies (passive/active): seeking, avoiding, sensitivity, and low registration. Each quadrant is classified as much less than others, less than others, similar to others, more than others, or much more than others. The AASP consists of 60 items rated on a 5-point Likert scale (1 = never to 5 = always). In this study, the Korean version of the AASP was used. Test–retest reliability coefficients were .81 for seeking, .88 for avoiding, .87 for sensitivity, and .91 for low registration (Jung, 2016).

### *D. Occupational Self Assessment*

The Occupational Self Assessment (OSA) is a self-report measure evaluating perceived competence and value in occupational performance and participation (Kielhofner et al., 2010). It consists of 21 items categorized into performance, habituation, and volition. The performance domain includes concentrating on my tasks, Physically doing what I need to do, Taking care of the place where I live, Taking care of myself, Taking care of others for whom I am responsible, Getting where I need to go, Managing my finances, Managing my basic needs (food, medicine), Expressing myself to others, Getting along with others, Identifying and solving problems, Relaxing and enjoying myself, Getting done what I need to do, Having a satisfying routine, Handling my responsibilities, Being involved as a student, worker, volunteer, and/or family member, Doing activities I like, Working towards my goals, Making decisions based on what I think is important, Accomplishing what I set out to do, Effectively using my abilities. Items are rated on a 4-point Likert scale from 1 (great difficulty / not important) to 4 (very well / most important). Rasch analysis showed item separation reliability of .88 and .80, and person separation reliability of .88 and .82 for competence and value, respectively (Popova et al., 2019).

### *E. Statistical Analyses*

Descriptive statistics were used to calculate the mean and standard deviation of each variable to identify general characteristics. Pearson correlation analysis was conducted to examine the associations among sensory processing patterns, SDNN, and interoception. Multiple regression analysis was performed to investigate how sensory processing patterns predicted occupational experience, with competence and value scores from the OSA set as dependent variables and quadrant scores of the AASP as independent variables. Statistical analyses were conducted using SPSS version 25.0 (Corp., Armonk, NY, USA), and the significance level was set at .05.

## III. RESULTS

### A. Participant Characteristics

This study was conducted with undergraduate students in early adulthood. Among the total of 43 participants, 35 (81.4%) were female. Table 1 presents the mean and standard deviation of SDNN, Heartbeat Error Score, AASP quadrants, and OSA competence and value.

TABLE I. GENERAL CHARACTERISTICS OF PARTICIPANTS ($N$=43)

| Variable | | *M*±*SD* |
|---|---|---|
| Age | | 20.51±1.737 |
| SDNN(ms) | | 48.58±17.77 |
| Heartbeat Error Score | | .32±.16 |
| AASP | Seeking | 36.33±7.22 |
| | Avoiding | 39.91±9.91 |
| | Sensitivity | 39.02±8.65 |
| | Low Registration | 32.07±8.15 |
| OSA | Competence | 58.51±7.03 |
| | Value | 57.37±9.32 |

### B. Correlation between Sensory Processing, SDNN, and Interoception

Low registration among sensory processing patterns showed a negative correlation with SDNN ($r$=−.317*) and a positive correlation with the Heartbeat Error Score ($r$=.343*) (Table 2). These results indicate that higher low registration scores were associated with lower HRV and greater error in heartbeat perception.

TABLE 2. CORRELATIONS BETWEEN SENSORY PROCESSING, SDNN AND HEARTBEAT ERROR SCORE

| AASP (*n*=43) | SDNN (*n*=43) | Heartbeat Error Score (*n*=34) |
|---|---|---|
| Seeking | .088 | .249 |
| Avoiding | -.260 | .133 |
| Sensitivity | -.297 | .133 |
| Low registration | -.317* | .343* |

AASP: Adolescent/Adult Sensory Profile, SDNN: Standard Deviation of Normal-to-Normal intervals
*$p$<.05

### C. Effects of Sensory Processing Patterns on OSA

Low registration among sensory processing patterns significantly predicted occupational competence ($\beta$=−.448, $p$=.015) (Table 3). This indicates that higher low registration scores were associated with lower perceived competence in daily occupational performance. In contrast, no sensory processing patterns significantly predicted occupational value.

TABLE 3. EFFECT OF SENSORY PROCESSING PATTERNS ON OSA. ($N$=43)

| OSA | AASP | *β* | *t* | *p* |
|---|---|---|---|---|
| Competence | Seeking | -.072 | -.503 | .618 |
| | Avoiding | -.093 | -.473 | .639 |
| | Sensitivity | -.087 | -.387 | .709 |
| | Low Registration | -.448 | -2.559 | .015 |
| | Adjusted $R^2$=.248, F[4, 38]=4.462, $p$<.05* | | | |
| Value | Seeking | -.009 | -.053 | .958 |
| | Avoiding | .189 | .797 | .430 |
| | Sensitivity | -.128 | -.476 | .637 |
| | Low Registration | -.001 | -.003 | .998 |
| | Adjusted $R^2$=-.084, F[4, 38]=.188, $p$=.943 | | | |

AASP: Adolescent/Adult Sensory Profile, OSA: Occupational Self Assessment
*$p$<.05

## IV. DISCUSSION

This study aimed to examine the relationships among sensory processing patterns, autonomic nervous system function, and occupations in young adults. The results showed that among sensory processing patterns, low registration had a significant negative correlation with SDNN, an index of heart rate variability, and a positive correlation with the Heartbeat Error Score. In addition, low registration significantly predicted competence among the subdomains of occupations, but did not have a significant effect on value. These findings suggest that low registration may be closely related to autonomic regulation, interoception, and subjective competence in occupations.

In early adulthood, higher scores of low registration in sensory processing patterns were significantly associated with decreased heart rate variability and interoception. High low-registration scores are characterized by high neurological thresholds and passive self-regulation, reflecting reduced sensitivity to sensory stimuli (Dunn, 2007). This finding can be interpreted within the framework of Porges (2007) polyvagal theory, which posits that the autonomic nervous system determines survival strategies based on sensory information from both the external environment and the internal body, and that the regulation of visceral states and maintenance of homeostasis presuppose a feedback system consisting of motor, sensory, and regulatory components. Indeed, previous studies have reported significant associations between sensory processing patterns and autonomic nervous system activation and reactivity (Kerley et al., 2022), and decreased parasympathetic activity has been linked to specific sensory processing patterns (Hattori et al., 2023). Therefore, the low responsiveness to sensory input observed in individuals with high low-registration traits may interfere with the recognition and processing of sensory information necessary for autonomic regulation, thereby potentially reducing awareness of internal physiological states and the ability to regulate them.

Low registration in sensory processing was found to have a significant effect on occupational competence. Individuals with a low-registration pattern have difficulty perceiving sensory events in daily life that others readily notice (Dunn, 2001). This suggests that individuals with higher low-registration scores may have difficulty adequately perceiving sensory feedback during occupational performance, which may ultimately lead to a reduced sense of occupational competence. This finding supports previous research indicating that sensory processing patterns vary across individuals and may influence activity engagement, community participation, and patterns of time use in daily life (Bagatell et al., 2022).

This study has some limitations that suggest directions for future research. Although this study examined the relationships among sensory processing patterns, autonomic nervous system

function, and occupations, the relatively small sample size, particularly in the Heartbeat Tracking Task analysis, should be considered when interpreting the findings. Moreover, as the study was conducted within a limited age range, research involving participants from a broader age range may help clarify the applicability of the present findings. In addition, autonomic nervous system function was examined using SDNN, a time-domain index of heart rate variability. Future research incorporating additional physiological indicators, such as frequency-domain measures, may provide a more comprehensive understanding of autonomic regulation in relation to sensory processing characteristics. Finally, given the cross-sectional design of this study, the observed relationships among sensory processing patterns, autonomic regulation, and occupations should be interpreted with caution. Accordingly, research using complementary study designs may contribute to a deeper understanding of these associations.

## V. CONCLUSIONS

This study examined the relationships among sensory processing patterns, autonomic nervous system function, and occupations in young adults. The results showed that low registration was associated with autonomic regulation and interoception and significantly predicted occupational competence. These findings suggest that sensory processing characteristics may be related to physiological regulation and subjective perception of occupations. Therefore, future research is needed to expand the understanding of the relationships among sensory processing patterns, autonomic nervous system, and occupation across diverse populations.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176).

## REFERENCES

[1] Bagatell, N., Chan, D. V., Syu, Y.-C., Lamarche, E. M., & Klinger, L. G. (2022). Sensory processing and community participation in autistic adults. *Frontiers in Psychology, 13*, 876127. https://doi.org/10.3389/fpsyg.2022.876127

[2] Bayram-Weston, Z., Andrade-Sienz, M., & Knight, J. (2022). Nervous system 6: The autonomic nervous system – anatomy and function. *Nursing Times, 118*(8).

[3] Brown, C., Tollefson, N., Dunn, W., Cromwell, R., & Filion, D. (2001). The Adult Sensory Profile: Measuring patterns of sensory processing. *American Journal of Occupational Therapy, 55*(1), 75–82. https://doi.org/10.5014/ajot.55.1.75

[4] Brown, C., & Dunn, W. (2002). *Adolescent/Adult Sensory Profile: User's manual*. San Antonio, TX: Psychological Corporation.

[5] Dunn, W. (2001). The sensations of everyday life: Empirical, theoretical, and pragmatic considerations. *American Journal of Occupational Therapy, 55*(6), 608–620. https://doi.org/10.5014/ajot.55.6.608

[6] Dunn, W., Saiter, J., & Rinner, L. (2002). Asperger syndrome and sensory processing: A conceptual model and guidance for intervention planning. *Focus on Autism and Other Developmental Disabilities, 17*(3), 172–185. https://doi.org/10.1177/10883576020170030701

[7] Dunn, W. (2007). Supporting children to participate successfully in everyday life by using sensory processing knowledge. *Infants & Young Children, 20*(2), 84–101. https://doi.org/10.1097/01.IYC.0000264477.05076.5d

[8] Eeles, A. L., Spittle, A. J., Anderson, P. J., Brown, N., Lee, K. J., Boyd, R. N., & Doyle, L. W. (2013). Assessments of sensory processing in infants: A systematic review. *Developmental Medicine & Child Neurology, 55*(4), 314–326. https://doi.org/10.1111/j.1469-8749.2012.04434.x

[9] Grippo, A. J., & Johnson, A. K. (2002). Biological mechanisms in the relationship between depression and heart disease. *Neuroscience & Biobehavioral Reviews, 26*(8), 941–962. https://doi.org/10.1016/S0149-7634(03)00003-4

[10] Harrold, A., Keating, K., Larkin, F., & Setti, A. (2024). The association between sensory processing and stress in the adult population: A systematic review. *Applied Psychology: Health and Well-Being, 16*(4), 2536–2566. https://doi.org/10.1111/aphw.12554

[11] Hattori, R., Irie, K., Mori, T., Tsurumi, K., Murai, T., & Inadomi, H. (2023). Sensory processing, autonomic nervous function, and social participation in people with mental illnesses. *Hong Kong Journal of Occupational Therapy, 36*(1), 39–47. https://doi.org/10.1177/15691861231177355

[12] Jung, H. R. (2016). *Technical adequacy of Korean Adolescent/Adult Sensory Profile (K-AASP)* [Doctoral dissertation, Inje University].

[13] Kerley, L., Meredith, P., & Harnett, P. (2022). Investigating autonomic biomarkers of sensory processing patterns in young adults. *British Journal of Occupational Therapy, 85*(11), 900–909. https://doi.org/10.1177/03080226221089845

[14] Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., Feusner, J. D., Garfinkel, S. N., Lane, R. D., Mehling, W. E., Meuret, A. E., Nemeroff, C. B., Oppenheimer, S., Petzschner, F. H., Pollatos, O., Rhudy, J. L., Schramm, L. P., Simmons, W. K., Stein, M. B., ... Paulus, M. P., & the Interoception Summit 2016 participants. (2018). Interoception and mental health: A roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(6), 501–513. https://doi.org/10.1016/j.bpsc.2017.12.004

[15] Kielhofner, G., Dobria, L., Forsyth, K., & Kramer, J. (2010). The Occupational Self Assessment: Stability and the ability to detect change over time. *OTJR: Occupation, Participation and Health, 30*(1), 11–19. https://doi.org/10.3928/15394492-20091214-03

[16] Kilroy, E., Aziz-Zadeh, L., & Cermak, S. (2019). Ayres theories of autism and sensory integration revisited: What contemporary neuroscience has to say. *Brain Sciences, 9*(3), 68. https://doi.org/10.3390/brainsci9030068

[17] Lischke, A., Pahnke, R., Mau-Moeller, A., & Weippert, M. (2021). Heart rate variability modulates interoceptive accuracy. *Frontiers in Neuroscience, 14*, 612445. https://doi.org/10.3389/fnins.2020.612445

[18] Popova, E. S., Ostrowski, R. K., Wescott, J., & Taylor, R. R. (2019). Development and validation of the Occupational Self-Assessment–Short Form (OSA–SF). *American Journal of Occupational Therapy, 73*(3), 7303205020p1–7303205020p10. https://doi.org/10.5014/ajot.2019.030288

[19] Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology, 74*(2), 116–143. https://doi.org/10.1016/j.biopsycho.2006.06.009

[20] Quigley, K. S., Kanoski, S., Grill, W. M., Barrett, L. F., & Tsakiris, M. (2021). Functions of interoception: From energy regulation to experience of the self. *Trends in Neurosciences, 44*(1), 29–38. https://doi.org/10.1016/j.tins.2020.09.008

[21] Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology, 18*(4), 483–488. https://doi.org/10.1111/j.1469-8986.1981.tb02486.x

[22] Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health, 5*, 258. https://doi.org/10.3389/fpubh.2017.00258

[23] Task Force of the European Society of Cardiology, & the North American Society of Pacing and Electrophysiology. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal, 17*(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

[24] Tsakiris, M., & Critchley, H. (2016). Interoception beyond homeostasis: Affect, cognition and mental health. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1708), 20160002. https://doi.org/10.1098/rstb.2016.0002

# A Systematic Review of Virtual Reality (VR) -based Cognitive Interventions for Children with Attention-Deficit Hyperactivity Disorder (ADHD)

Yaena Ha[1], Deukgeun Yoon[2], and Eun Young Kim[3,*]
[1]*Emotional and Intelligent Child Care System Convergence Research Center, Soonchunhyang University, Republic of Korea*
[2]*Department of Occupational Therapy, Graduate School, Soonchunhyang University, Republic of Korea*
[3]*Department of Occupational Therapy, Soonchunhyang University, Republic of Korea*
*Contact: eykim@sch.ac.kr

*Abstract*— **The purpose of this study is to systematically review randomized controlled trials (RCTs) regarding virtual reality (VR)-based cognitive interventions for children with Attention-Deficit Hyperactivity Disorder (ADHD). A search was conducted for studies published over the past decade (2015–2025) focusing on VR-based cognitive interventions for this population. The search strategy involved combining keywords such as children with ADHD, digital devices with VR, intervention, cognition, and its specific domains. Databases including PubMed, EMBASE, Web of Science, and CINAHL Complete were utilized. Out of 770 identified articles, four studies were finally selected for analysis according to inclusion and exclusion criteria based on PRISMA guidelines. VR-based cognitive interventions for children with ADHD showed positive effects in cognitive areas such as Attention (omissions, commissions, detectability, and visual search), Memory (memory, working memory, and material organization), and Executive functions (inhibition, set shifting, and multitasking). Beyond cognition, positive outcomes were observed in ADHD symptoms, social skills, self-control, initiative, emotional control, internalizing, sleep, and hemoencephalographic (HEG) measures.**

## I. INTRODUCTION

ADHD is a neurodevelopmental disorder characterized by inattention, hyperactivity, and impulsivity [1]. According to the latest NSCH data, the prevalence of ADHD diagnosis among U.S. children and adolescents has reached 11.4% [2], [3].

Due to the core symptoms of ADHD, children and adolescents experience academic, social, and emotional difficulties, which present continuous challenges to their school adjustment [4], [5]. Difficulties encountered in childhood and adolescence due to ADHD tend to have long-term consequences that extend into adult life [6] – [10]. Consequently, emphasis should be placed on providing ongoing monitoring and interventions for children with ADHD from their early developmental stages [7], [10].

A proportion of children diagnosed with ADHD experience learning/cognitive disorder [4], [11]. A comparison of cognitive functions between children with ADHD and typically developing (TD) children revealed that the ADHD group showed significant cognitive delays compared to their TD peers across domains, including verbal comprehension (similarities, vocabulary), perceptual reasoning (block design, picture concepts, matrix reasoning), working memory (digit span, letter-number sequencing), and processing speed (coding, symbol search) [11]. Interventions to improve cognitive difficulties in children with ADHD include VR-based cognitive training, multiple cognitive training, and working memory training [12] – [14]. These cognitive interventions are effective in enhancing the cognitive functions of children with ADHD [12] – [15].

VR-based interventions are drawing significant attention as an innovative tool in modern rehabilitation [16], [17]. VR-based cognitive interventions offer several distinct advantages over traditional methods. First, in terms of therapeutic efficacy, they enable precise improvements in core cognitive processes—such as working memory, executive function, and attention—by leveraging superior ecological validity. Second, regarding operational benefits, these interventions provide a highly flexible environment tailored to the individual needs of children with ADHD, while ensuring a controlled training space that filters out distractions and guarantees safety from real-world hazards. Furthermore, VR fosters high levels of engagement, which maximizes treatment motivation and enhances both time and cost efficiency. Finally, from an assessment perspective, the provision of objective behavioral data empowers therapists to perform more accurate diagnoses and implement systematic, data-driven rehabilitation [18].

While existing systematic reviews have demonstrated the technical feasibility and long-term efficacy of VR-based interventions, discussions regarding standardized outcome measures and personalized protocols applicable to actual clinical settings remain in their infancy. In particular, previous studies have faced limitations in ensuring consistency of results by pooling research with heterogeneous designs [16] – [20]. Therefore, the purpose of this study is to systematically RCTs on VR-based cognitive interventions for children with ADHD to provide evidence base.

## II. METHOD

A search was conducted for studies published between 2015 and 2024 regarding VR-based cognitive interventions for children with ADHD. The search strategy involved combining keywords such as children with ADHD, digital devices including VR, interventions, cognition, and its specific domains. PubMed, EMBASE, Web of Science, and CINAHL Complete were utilized as databases for the search.

A total of 770 records were identified through database searches, comprising 144 from PubMed, 141 from EMBASE, 470 from Web of Science, and 15 from CINAHL Complete. These records were then screened based on the inclusion and exclusion criteria of the PRISMA guidelines [21]. After removing 90 duplicate records, the remaining 680 studies were screened by title and abstract. Based on the inclusion and exclusion criteria, studies were excluded if they did not involve children or adolescents with ADHD (n = 93), utilized non-digital interventions (n = 332), were not RCTs (n = 67), or involved pharmacological interventions (n = 129). The full texts of 59 studies were then assessed for eligibility, and ultimately, 4 RCTs studies utilizing VR as an intervention tool—excluding general digital interventions such as PC or smartphone-based programs—were included in the final analysis.

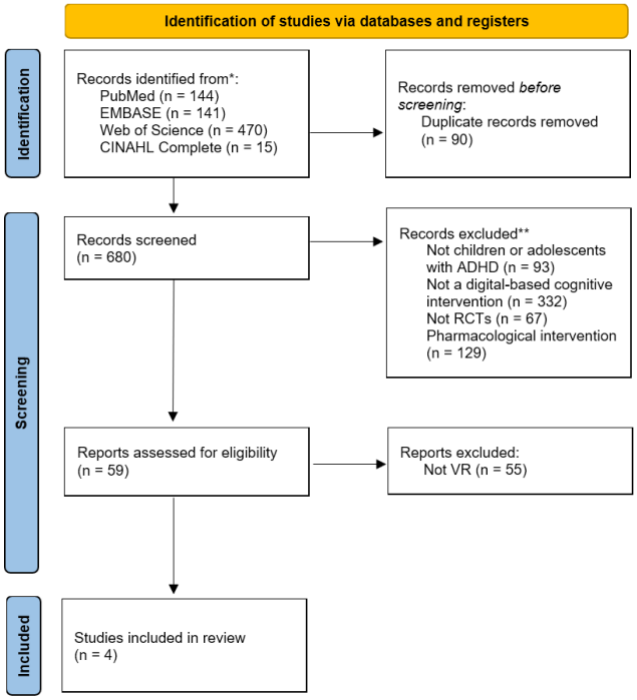



Fig. 1 PRISMA flow chart

## III. RESULTS

### A. Trends and Evidence Level of Analyzed Studies

Four studies were published, with one study appearing in each of the years 2020, 2021, 2024, and 2025. All analyzed studies were RCTs, representing the high level of evidence [22].

### B. General Characteristics of Research Participants

The research participants consisted entirely of children with ADHD, with a total sample size of 304. The age range spanned from 6 to 18 years, with a weighted mean age of 10.98 years.

TABLE I
OVERVIEW OF STUDY TRENDS AND PARTICIPANT CHARACTERISTICS

| Author | Year | N (M/F) | Mean age (range) |
|---|---|---|---|
| Martin-Moratinos et al. [23] | 2025 | 76 (61/15) | 12.68y (7-18y) |
| Wong et al. [24] | 2024 | 90 (73/17) | 8.53y (6-12 y) |
| Skalski et al. [25] | 2021 | 90 (74/13) | 12.75y (9-15y) |
| Tabrizi et al. [17] | 2020 | 48 | 9.58y (8-12y) |

### C. Effectiveness of VR-Based Cognitive Interventions

In this study, classification of the main cognitive processes was categorized into three areas—Attention, Memory, and Executive functions—based on the research by Schmidt et al. [26]. Following this classification, VR-based cognitive interventions for children with ADHD demonstrated positive effects across all three areas: Attention (omissions, commissions, detectability, and visual search), Memory (memory, working memory, and material organization), and Executive functions (inhibition, set shifting, and multitasking).

In extra-cognitive areas, positive effects were also observed across several functional areas: Social (social skills and prosocial scale), Emotional (emotional symptoms, emotional control, initiative, and self-control), Behavioral (internalizing), Sleep, and Physiological (HEG measures).

TABLE IIIII
EFFECTIVENESS OF VR-BASED COGNITIVE INTERVENTIONS

| Area | | Outcome | n |
|---|---|---|---|
| Cognitive | Attention | Visual Search | 1 |
| | | Detectability | 1 |
| | | Omissions | 1 |
| | | Comissions | 2 |
| | Memory | Memory | 1 |
| | | Working memory | 1 |
| | | Material organization | 1 |
| | Executive functions | Inhibition | 1 |
| | | Set shifting | 1 |
| | | Multitasking | 1 |
| Extra-cognitive | Social | Social skills | 1 |
| | | Prosocial scale | 1 |
| | Emotional | Emotional symptoms | 1 |
| | | Emotional control | 1 |
| | | Initiative | 1 |
| | | Self-control | 1 |
| | Behavioral | Internalizing | 1 |
| | Sleep | Sleep | 1 |
| | Physiological | HEG | 1 |

### D. Assessment Tools for Measuring Effects

The following assessment tools were used to compare cognitive functions before and after the intervention: BRIEF-2 (behavior rating inventory of executive function), BRIEF-Parents, CTMT-2 (comprehensive trail making test), CPT-3

(conners continuous performance test), the short form of the mackworth clock task, the visual search task, the multitasking test, raven's colored progressive matrices, raven's progressive matrices test, digit span subscale (wechsler IQ test), and corsi block-tapping test.

TABLE IVVVI
ASSESSMENT TOOLS FOR MEASURING EFFECTS

| Area | Assessment | | | n |
|---|---|---|---|---|
| Cognitive | BRIEF | | -2 | 2 |
| | | | -Parents | |
| | CTMT-2 | | | 1 |
| | CPT-3 | | | 1 |
| | The Mackworth Clock Task | | | 1 |
| | The Visual Search Task | | | 1 |
| | The Multitasking Test | | | 1 |
| | Raven's Colored Progressive Matrices | | | 1 |
| | Raven's Progressive Matrices Test | | | 1 |
| | Digit Span Subscale (Wechsler IQ Test) | | | 1 |
| | Corsi Block-Tapping Test | | | 1 |
| Extra-cognitive | ADHD symptom | SNAP-IV | | 1 |
| | | CPRS | | 1 |
| | | CGI | | 1 |
| | Social | SSIS | -Rating scales | 2 |
| | | | -Parents | |
| | | SIS | | 1 |
| | Emotional | SDQ | | 1 |
| | Sleep | SDSC | | 1 |
| | Physiological | HEG system (Medi-TECH Electronic GmbH) | | 1 |
| | Safety | SSQ | | 1 |
| | | GASA | | 1 |
| | | UKU | | 1 |

BRIEF = Behavior Rating Inventory of Executive Function, CGI = Clinical Global Impression, CPRS = Conners Abbreviated Symptom Questionnaire, CPT = Conners Continuous Performance Test, CTMT = Comprehensive Trail Making Test, GASA = The Game Addiction Scale for Adolescents, SDQ = Strengths and Difficulties Questionnaire, SDSC = Sleep Disturbance Scale for Children, SNAP = Swanson, Nolan, and Pelham Rating Scale, SSI =The clinical psychologist used a modified version of Riggio basic social skills assessment, SSIS = The Social Skills Improvement System, SSQ = The Simulator Sickness Questionnaire, UKU = The questionnaire Udvalg für Kliniske Undersogelser

## IV. DISCUSSIONS

This study systematically reviewed the effectiveness of VR-based cognitive interventions for children with ADHD through RCTs published over the past decade. The analysis revealed that these interventions have a positive impact on core cognitive processes, categorized into three primary areas: Attention, including improvements in omissions, commissions, detectability, and visual search; Memory, encompassing memory, working memory, and material organization; and Executive functions, specifically in the domains of inhibition, set shifting, and multitasking. These advancements highlight the comprehensive potential of VR-based interventions in addressing the multifaceted cognitive deficits of ADHD.

The reviewed research analyzed in this study encompass a broad spectrum of school-aged children and adolescents with ADHD, ranging from 6 to 18 years, suggesting the potential for VR interventions to be widely applied across various developmental stages. A notable limitation of this research is that it did not separately analyze the effects on children and adolescents. While the reviewed studies spanned a wide age range (6–18 years), the lack of sub-group analysis makes it difficult to determine whether the efficacy of VR interventions differs significantly between early childhood and adolescence. Given that brain development stages and the specific characteristics of cognitive deficits differ significantly between childhood and adolescence [27], [28], future efforts should prioritize the development of customized VR cognitive intervention contents tailored to the developmental profiles of younger children with ADHD, alongside age-specific efficacy validations.

The analysis demonstrates that VR-based interventions yield precise and multifaceted effects in addressing cognitive deficits in children with ADHD. Specifically, the reduction in omissions and commissions, alongside enhanced inhibition, underscores that virtual environments are optimized tools for training selective attention and impulse control [16], [29]. Furthermore, the positive shifts in visual search, working memory, and set shifting suggest that the highly immersive stimuli inherent in VR facilitate complex information-processing capabilities [25], [30] − [32]. The positive outcomes in material organization and multitasking suggest that VR-based gains may support high-level executive functions relevant to daily-life performance [31], [33]. Consequently, the development of future cognitive intervention contents must adopt these specific cognitive indicators as core design modules, implementing a graded cognitive-load algorithm that begins with foundational attention training and progressively stimulates complex executive functions.

The findings of this study identify the BRIEF as the most pivotal assessment tool for verifying cognitive improvement in children with ADHD. In particular, the BRIEF exhibits superior clinical utility compared to other instruments due to its multidimensional capacity to evaluate the complex cognitive deficits inherent in ADHD, such as inhibition, shifting, emotional control, working memory, and planning/organization [32]. These results suggest that future cognitive intervention contents must transcend simple functional training by adopting the subscales of the BRIEF as core design indicators. Furthermore, the VR-based cognitive interventions analyzed in this study demonstrated therapeutic efficacy while employing a safety management framework incorporating the SSQ and GASA. These findings support that potential side effects, such as cybersickness or addiction risks, can be effectively prevented under appropriate control protocols, providing a critical foundation for enhancing the technological acceptance of future cognitive intervention contents [18], [19]. Therefore, sophisticated intervention scenarios must be constructed to ensure that gains achieved within virtual environments—validated by both cognitive efficacy and clinical safety—directly translate into tangible improvements in behavioral regulation as observable in real-world settings.

The primary limitation of this study is the very small number of papers included in the final analysis, despite covering a 10-year period. Furthermore, a significant limitation of this study is the absence of a differentiated analysis between children and adolescents. Additionally,

despite the use of strict RCT designs, significant variations in clinical environments, assessment protocols, and intervention strategies—including the types of VR devices used (HMD vs. Screen-based) pose a challenge to the external validity and generalizability of the results, making it difficult to determine the superiority of any specific intervention model.

## V. CONCLUSIONS

This study showed that VR-based cognitive interventions effectively enhance both core executive functions and extra-cognitive domains in children with ADHD, emphasizing that future development must adopt BRIEF subscales as key performance indicator and embed standardized safety protocols to establish an evidence-based digital therapeutic (DTx) model that maximizes real-world behavioral generalization.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176)

## REFERENCES

[1] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders,* 5th ed. Arlington, VA, USA: American Psychiatric Publishing, 2013. doi: 10.1176/appi.books.9780890425596.

[2] M. L. Danielson et al., "ADHD Prevalence Among U.S. Children and Adolescents in 2022: Diagnosis, Severity, Co-Occurring Disorders, and Treatment," *J. Clin. Child Adolesc. Psychol.*, vol. 53, no. 3, pp. 343–360, 2024, doi: 10.1080/15374416.2024.2335625.

[3] B. Zablotsky *et al.*, "Prevalence and trends of developmental disabilities among children in the United States: 2009–2017," *Pediatrics*, vol. 144, no. 4, Oct. 2019, Art. no. e20190851, doi: 10.1542/peds.2019-0851.

[4] G. J. DuPaul, A. Chronis-Tuscano, M. L. Danielson, and S. N. Visser, "Predictors of receipt of school services in a national sample of youth with ADHD," *Journal of Attention Disorders*, vol. 23, no. 11, pp. 1303-1319, Sep. 2019.

[5] G. Giannakopoulos, "Adolescents with ADHD in the school environment: A comprehensive review of academic, social, and emotional challenges and interventions," *J. Clin. Images Med. Case Rep.*, vol. 6, no. 3, Art. no. 3528, 2025, doi: 10.52768/2766-7820/3528.

[6] R. Di Lorenzo *et al.*, "Children and adolescents with ADHD followed up to adulthood: A systematic review of long-term outcomes," *Acta Neuropsychiatr.*, vol. 33, no. 6, pp. 283–298, Dec. 2021.

[7] R. G. Klein *et al.*, "Clinical and functional outcome of childhood attention-deficit/hyperactivity disorder 33 years later," *Arch. Gen. Psychiatry*, vol. 69, no. 12, pp. 1295–1303, Dec. 2012.

[8] A. Caye, A. V. Spadini, R. G. Karam, *et al.*, "Predictors of persistence of ADHD into adulthood: a systematic review of the literature and meta-analysis," *Eur. Child Adolesc. Psychiatry*, vol. 25, no. 11, pp. 1151–1159, Nov. 2016, doi: 10.1007/s00787-016-0831-8.

[9] S. Mannuzza and R. G. Klein, "Long-term prognosis in attention-deficit/hyperactivity disorder," *Child Adolesc. Psychiatr. Clin. N. Am.*, vol. 9, no. 3, pp. 711–726, Jul. 2000.

[10] L. Kern *et al.*, "Multisetting assessment-based intervention for young children at risk for attention deficit hyperactivity disorder: Initial effects on academic and behavioral functioning," *Sch. Psychol. Rev.*, vol. 36, no. 2, pp. 237–255, Jun. 2007.

[11] D. Ünal *et al.*, "Comparative analysis of the WISC-IV in a clinical setting: ADHD vs. non-ADHD," *Arch. Pédiatr.*, vol. 28, no. 1, pp. 16–22, Jan. 2021.

[12] S. Bioulac *et al.*, "Virtual remediation versus methylphenidate to improve distractibility in children with ADHD: a controlled randomized clinical trial study," *J. Atten. Disord.*, vol. 24, no. 2, pp. 326–335, Jan. 2020, doi: 10.1177/1087054718772486.

[13] M. van Dongen-Boomsma, M. A. Vollebregt, J. K. Buitelaar, and D. Slaats-Willemse, "Working memory training in young children with ADHD: A randomized placebo-controlled trial," *J. Child Psychol. Psychiatry*, vol. 55, no. 8, pp. 886–896, Aug. 2014, doi: 10.1111/jcpp.12218.

[14] S. Chen *et al.*, "Which factor is more relevant to the effectiveness of the cognitive intervention? A meta-analysis of randomized controlled trials of cognitive training on symptoms and executive function behaviors of children with attention deficit hyperactivity disorder," *Front. Psychol.*, vol. 12, Art. no. 810298, Jan. 2022, doi: 10.3389/fpsyg.2021.810298.

[15] A. Veloso, S. G. Vicente, and M. G. Filipe, "Effectiveness of cognitive training for school-aged children and adolescents with attention deficit/hyperactivity disorder: a systematic review," *Front. Psychol.*, vol. 10, Art. no. 2983, Jan. 2020, doi: 10.3389/fpsyg.2019.02983.

[16] D. Romero-Ayuso *et al.*, "Effectiveness of virtual reality-based interventions for children and adolescents with ADHD: A systematic review and meta-analysis," *Children*, vol. 8, no. 2, Art. no. 70, Feb. 2021, doi: 10.3390/children8020070.

[17] M. Tabrizi, G. Manshaee, A. Ghamarani, J. Rasti, *Int. Arch. Health Sci.* 7 (2020) 37

[18] A. Bashiri, M. Ghazisaeedi, and L. Shahmoradi, "The opportunities of VR in the rehabilitation of children with ADHD: a literature review," *Korean J. Pediatr.*, vol. 60, no. 11, p. 337, 2017

[19] Y. Cao *et al.*, "Virtual reality interventions for attention deficit/hyperactivity disorder: a systematic review," *npj Digit. Med.*, vol. 9, Art. no. 42, Mar. 2026, doi: 10.1038/s41746-026-00123-x.

[20] N. Corrigan, C. R. Păsărelu, and A. Voinescu, "Immersive virtual reality for improving cognitive deficits in children with ADHD: a systematic review and meta-analysis," *Virtual Real.*, vol. 27, no. 4, pp. 3545–3564, Dec. 2023, doi: 10.1007/s10055-023-00813-w.

[21] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, Art. no. n71, Mar. 2021, doi: 10.1136/bmj.n71.

[22] A. L. Rosner, "Evidence-based medicine: Revisiting the pyramid of priorities," *J. Bodyw. Mov. Ther.*, vol. 16, no. 1, pp. 42–49, Jan. 2012, doi: 10.1016/j.jbmt.2011.05.003.

[23] M. Martin-Moratinos, M. Bella-Fernández, M. Rodrigo-Yanguas, C. González-Tardón, C. Li, P. Wang, A. Royuela, P. Lopez-Garcia, H. Blasco-Fontecilla, *JMIR Serious Games* 13 (2025) e59124

[24] K. P. Wong, B. H. Zhang, C. Y. Y. Lai, Y. J. Xie, Y. Li, C. Li, J. Qin, *JMIR Serious Games* 12 (2024) e58963

[25] S. Skalski, K. Konaszewski, G. Pochwatko, R. Balas, J. Surzykiewicz, *Int. J. Psychophysiol.* 170 (2021) 59

[26] C. Schmidt, F. Collette, C. Cajochen, and P. Peigneux, "A time to think: circadian rhythms in human cognition," *Cogn. Neuropsychol.*, vol. 24, no. 7, pp. 755-789, Oct. 2007, doi: 10.1080/02643290701754158.

[27] I. Pascual Zapatero, R. Jurado Barba, and L. Esteban Rodríguez, "Cognitive profiles and developmental variations in ADHD: A comparative analysis of childhood and adolescent diagnoses," *Appl. Neuropsychol. Child*, pp. 1–8, 2024.

[28] Y. Qian, L. Shuai, R. C. Chan, Q. J. Qian, and Y. Wang, "The developmental trajectories of executive function of children and adolescents with attention deficit hyperactivity disorder," *Res. Dev. Disabil.*, vol. 34, no. 5, pp. 1434–1445, 2013.

[29] I. Pachiti, F. S. Milienos, T. C. Papadopoulos, and P. Dimitropoulou, "Child training in the Child ViReal Support Program: Combining iVR-based cognitive training and CBT techniques in a pilot study," *PLoS One*, vol. 21, no. 2, p. e0343364, 2026.

[30] T. D. Parsons, "Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences," *Front. Hum. Neurosci.*, vol. 9, p. 660, 2015.

[31] M. Gelsomini, N. La Porta, F. Fusco, and M. Papandrea, "Designing for Attention: Evaluating an Affordable Immersive VR Platform for Neurodiverse Children," in *Proc. 27th Int. ACM SIGACCESS Conf. Comput. Accessibility*, 2025, pp. 1–18.

[32] J. L. Sun, X. J. Chaw, S. Fresnoza, and H. I. Kuo, "Effects of virtual reality-based exercise intervention in young people with attention-deficit/hyperactivity disorder: a systematic review," *J. NeuroEng. Rehabil.*, vol. 22, no. 1, p. 139, 2025.

[33] L. A. Vinci, A. Passaro, and F. Stasolla, "Augmented, Virtual, and Mixed Reality Assessment and Training for Executive Functions in Children with ADHD: A Scoping Review," *Information*, vol. 17, no. 2, Art. no. 186, Feb. 2026, doi: 10.3390/info17020186.

[34] G. A. Gioia, P. K. Isquith, S. C. Guy, and L. Kenworthy, Behavior Rating Inventory of Executive Function, Second Edition (BRIEF-2): Professional Manual, 2nd ed. Lutz, FL, USA: PAR, 2015.

# Can a VLM's In-Context Learning Predict Video Saliency?

J. Ryu[1], Y. Park[2], and S. Seo[2,*]
[1]*Department of Applied Art and Technology, Chung-Ang University, Anseong, 17546, South Korea*
[2]*School of Art and Technology, Chung-Ang University, Anseong, 17546, South Korea*
* Contact: sanghyun@cau.ac.kr, phone +82-10 7273 0318

***Abstract*— This paper investigates the applicability of vision-language models (VLMs) to perform video saliency prediction (VSP) via in-context learning (ICL). Ground-truth saliency maps are converted into bounding box representations using a DBSCAN-based clustering algorithm, enabling structured ICL examples and proper outputs of VLM. Experiments on the Hollywood-2 dataset show that in-context examples improve prediction over the zero-shot baseline. However, using intersection over union (IoU) as quantitative evaluation metric, the values are consistently low across all conditions. This suggests that in-context learning alone is insufficient for reliable video saliency prediction and integration with dedicated vision models will be necessary to achieve reliable performance.**

## I. Introduction

Large language models (LLMs) have achieved remarkable results, spanning from simple text-based tasks to complex multimodal tasks. Building upon this foundation, vision-language models (VLMs) have extended these capabilities to the visual domain, enabling a wide range of understanding and generation tasks.

However, visual understanding and generation tasks extend beyond semantic understanding. These tasks require precise reasoning over spatial and temporal information (especially in the video domain). Such spatial reasoning presents a fundamental challenge for models whose outputs are primarily text-based, as raw textual descriptions are inherently limited in their ability to capture fine-grained spatial information.

The human visual system addresses this challenge effectively, directing attention to the most perceptually significant regions of a scene rapidly. Video saliency prediction (VSP) seeks to computationally model this capacity. It aims to estimate where human observers are likely to direct their gaze across video frames. Conventional VSP approaches typically rely on large datasets and task-specific model training, limiting their generalizability across diverse visual domains.

VLMs, by contrast, exhibit strong generalization capabilities through in-context learning (ICL), whereby a model adapts to a novel task given only a small number of demonstration examples, without any parameter updates. Recent work has further demonstrated that VLMs possess non-trivial visual grounding abilities, enabling them to associate natural language descriptions with specific spatial regions within an image or video [1].

For this reason, we study whether the VSP capability of a VLM can be elicited solely through ICL, without task-specific fine-tuning. Specifically, we formulate VSP as a structured prediction task in which the model is prompted to localize the most salient region in each video frame as a bounding box. Then, we examine how prediction quality varies as a function of the number of in-context examples (n-shots).

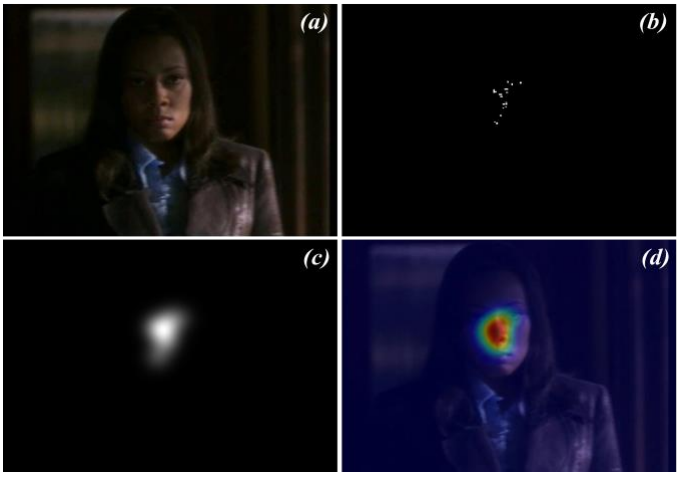



Fig. 1 Illustration of the VSP dataset workflow. (a) Input video frame. (b) Raw gaze fixation points collected via eye-tracking. (c) Smoothed saliency map generated by applying Gaussian blur to the fixation map. (d) Final saliency map overlaid on the input frame, representing the prediction target for VSP models.

Our contributions are the following: (1) We demonstrate that VLM-based VSP via ICL is potentially viable, achieving improved localization over the zero-shot baseline. (2) Enabling VLM to perform VSP, a process for converting heatmap-style saliency annotations into text-based bounding box representations was devised. (3) We provide an empirical analysis of the effect of in-context example quantity on VSP performance, revealing current VLM's limitation.

## II. Related Work

### A. VLM and In-Context Learning

Modern VLMs process visual and linguistic information in a unified manner by projecting tokenized images into the input space of language models, demonstrating strong performance across tasks including image captioning, VQA, and object detection. Since its introduction by [2], ICL has become a central paradigm in VLM research, enabling models to infer task-specific input-output patterns from a small number of in-prompt demonstrations without parameter updates. Recent studies have further shown that VLMs can leverage ICL to output bounding box coordinates for object localization [3].

### B. Video Saliency Prediction

VSP aims to predict the region most likely to attract human visual attention in each video frame, with eye-tracking-derived density heatmaps as the prediction target, as illustrated in Fig. 1. While recent transformer-based and

multimodal approaches have advanced in the field, they remain dependent on large-scale labelled datasets [4]. This work proposes an ICL-based approach as a training-free alternative to address this limitation.

## III. Methodology

We used Google's Gemini as the vision-language model and employed a video saliency prediction dataset to conduct experiments on VSP via in-context learning [5]. The overall pipeline is illustrated in Fig. 2.

### A. Model Selection

Google Gemini was selected as the VLM backbone due to its state-of-the-art performance among the vision-language models [6]. Specifically, we employed *gemini-3-flash-preview*, which supports an input token limit of approximately 1,000,000 and an output token limit of approximately 65,000 [7]. Given the scale
of the model, it is impossible to run model inference in a local computing environment. Therefore, all experiments were conducted using the Google AI Studio API.

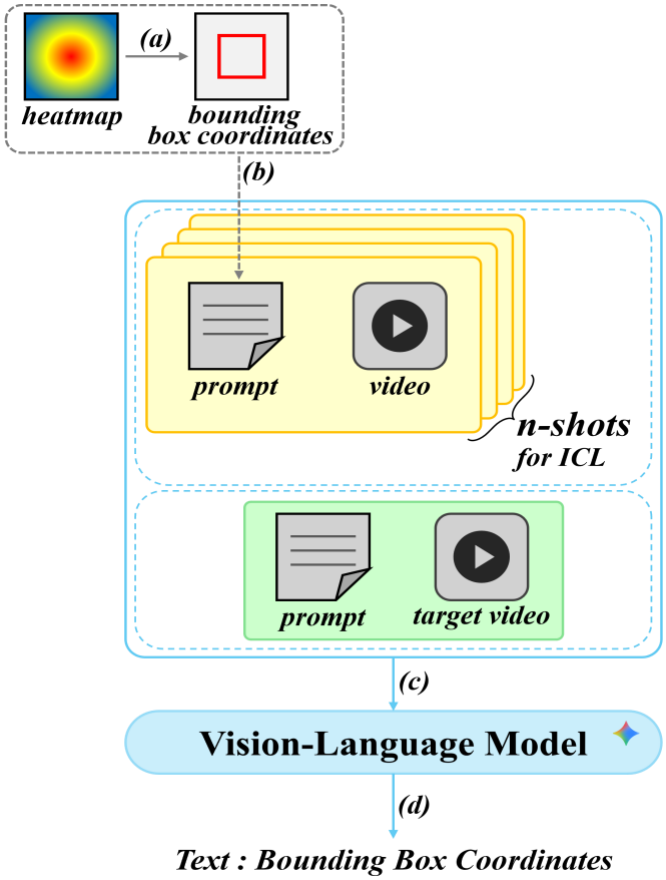



Fig. 2 Overall pipeline for VLM-based VSP via ICL. (a) Ground-truth saliency heatmaps are converted into bounding box coordinates to serve as a VLM-compatible representation. (b) The resulting coordinates are incorporated into text prompts paired with their corresponding video clips as in-context examples. (c) The VLM receives n-example pairs of prompt and video, followed by a single target pair consisting of a prompt and the target video without ground-truth coordinates. (d) The model outputs a sequence of per-frame bounding box coordinates representing the predicted salient region in each frame.

### B. Dataset

There are many open-source datasets for training and evaluating models [8]-[10]. Among these, the Hollywood-2 dataset was selected for its structural properties [11]. Unlike other datasets in which camera position, angle, or scene composition may vary in a single clip, Hollywood-2 is constructed by segmenting individual movie scenes into multiple consecutive shots. This characteristic minimizes confounding visual factors, enabling a more controlled evaluation of the VLM's VSP capability in isolation. In addition, we used only clips with a length of about three seconds (around 72 frames) in the experiments.

### C. Modified Output

The output of conventional VSP models consists of a set of per-frame heatmaps with pixel values in the range [0, 255]. Representing the full heatmap sequence as a token sequence would exceed the output limit of the VLM. To address this limit, the salient region in each frame is converted into the form of a bounding box, which serves both as the in-context learning example and as the target output format for the VLM.

The conversion procedure is as follows. For each frame of the ground-truth video saliency map, pixels below a predefined threshold are discarded. The remaining pixels are then clustered using the DBSCAN algorithm. Among the resulting clusters, only the cluster containing the pixel with the highest saliency value is remained. The minimum bounding box containing this cluster is computed, which is a compact and VLM-compatible representation of the salient region. This procedure was applied to all frames of all videos used in the experiments, and the resulting bounding boxes were saved as JSON files with video metadata. The entire workflow was implemented in Python using the scikit-learn library.

TABLE I
PROMPT TEMPLATES USED FOR EXPERIMENTS

| System Prompt |
|---|
| You are a visual saliency detector for video analysis.<br>**Task**: Identify the most visually salient region in each frame of the provided video.<br>**Analysis Requirements:**<br>- Process every frame sequentially from frame 1 to frame [total_frames]<br>- Identify regions where human visual attention is naturally drawn (based on contrast, motion, objects, composition)<br>- Select only one visually salient region<br>**Output Format:**<br>Raw coordinates only. **Each line = one frame.** No markdown, no tables.<br>**Important:**<br>- Process the full video duration. YOU MUST OUTPUT EXACTLY [total_frames] LINES<br>- If output is truncated due to length limits, stop at a complete frame and indicate continuation is needed |
| **Instruct Prompt** |
| Total frames : {frames}<br>Analyze all frames (1 to {frames}) and output accurate number base coordinates of bounding box. |

{frames} is a video-specific integer value.




Fig. 3 Qualitative VSP results across varying shot conditions.

### D. VSP via In-context Learning

Ten videos were randomly sampled from the dataset and designated as the fixed example set for in-context learning. Experiments were conducted from zero-shot to 10-shots inference, with the same set of examples used consistently in certain shot conditions. The system prompt describes the task, while the per-video input prompt includes the total frame count of the corresponding clip. The detailed prompt is written in Table I. Model outputs were parsed and serialized into JSON files, which are subsequently used for evaluations.

## IV. Experiments

Both qualitative and quantitative evaluations were conducted to test the proposed VLM-based VSP via ICL. Overall, providing in-context examples yields relatively better results than the zero-shot condition. Nevertheless, the results imply that ICL alone is insufficient to achieve reliable saliency prediction.

### A. Visual Comparison

Visual comparison is shown in Fig 3. The top row of the figure shows the ground-truth saliency heatmaps provided by the original dataset. It represents the per-frame distribution of human visual attention. Subsequent rows show prediction results under different shot conditions. The red bounding box indicates the ground-truth salient region converted from the heatmap, and the green bounding box is the region predicted by the model.

Providing bounding box coordinates alongside the target video yields more accurate prediction compared to the zero-shot condition. This suggests that in-context learning is effective for the VSP task.

TABLE II
Quantitative Evaluation - Mean IOU Scores

| N -shots | 0 | 2 | 4 | 6 | 8 | 10 |
|---|---|---|---|---|---|---|
| mean IOU | 0.1060 | 0.0529 | 0.1246 | 0.1458 | 0.0966 | 0.1449 |

### B. Quantitative Evaluation

Since VSP had been reformulated as bounding box prediction task, intersection over union (IoU) was adopted as the quantitative evaluation metric. For each video, the mean IoU was calculated by averaging the per-frame IoU scores across all frames. Subsequently, the per-video mean IoU scores were averaged within each shot condition to yield a score that represents each condition's performance.

Table II presents the mean IoU scores across varying numbers of in-context examples. The 6-shot condition achieved the highest mean IoU of 0.1458, suggesting that a moderate number of examples yields the most accurate predictions. Performance does not improve linearly with the number of shots; notably, the 8-shot condition records a mean IoU of 0.0966, underperforming relative to both the 6-shot and 10-shot conditions. This non-linear trend indicates that an excessive number of in-context examples does not necessarily lead to better results, which is likely attributable to the lost-in-the-middle phenomenon, in which models tend to underutilize information positioned in the middle of a long context window [12].

However, the mean IoU values remain low across all shot conditions, suggesting that in-context learning alone is insufficient for reliable VSP performance.

## V. Conclusions

In this work, we investigated the applicability of VLMs to the task of video saliency prediction via in-context learning. To bridge format style gap between heatmap-based saliency outputs and the text-based generation of VLMs, ground-truth saliency maps were converted into bounding box representations through DBSCAN-based clustering. This enables VLM to understand structured format for ICL example and return inference outputs.

Experimental results demonstrate that a number of in-context examples yields improved predictions over the zero-shot condition, confirming that VLMs can perform VSP via in-context learning. These findings indicate that ICL is a viable mechanism for eliciting spatial saliency reasoning from VLMs without any task-specific fine-tuning.

Nevertheless, the absolute IoU values observed across all conditions remain limited, indicating that ICL alone is insufficient to achieve reliable saliency localization. Integrating VLMs with dedicated computer vision models is necessary to achieve improved prediction performance, as suggested in [13].

Future work will explore additional datasets to assess the generalizability of the proposed approach, investigate principled selection strategies for ICL, and incorporate

established CV-based models as complementary components to further improve prediction accuracy.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176)

## REFERENCES

[1] T. Wei, S. Yanpeng, G. Qinying and Li, Zechao, "Visual position prompt for mllm based visual grounding," *IEEE Transactions on Multimedia(Early Access)*, pp. 1-16, Jan. 2026.

[2] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. K. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language Models are Few-Shot Learners," *Advances in Neural Information Processing Systems*, vol. 33, p.1877, 2024.

[3] S. Doveh, N. Shabtay, E. Schwartz, H. Kuehne, R. Giryes, R. Feris, L. Karlinsky, J. Glass, A. Arbelle, S. Ullman, and M. J. Mirza, "Teaching VLMs to Localize Specific Objects from In-context Examples," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9572-9582, 2025.

[4] X. Zhou, J. Lin, D. Chen, D. Liu, J. Zhang, and R. Cong, "Transformer-Based Multi-Scale Feature Integration Network for Video Saliency Prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 12, pp. 7696 - 7707, Mar. 22, 2023.

[5] (2026) The Google DeepMind website. [Online]. Available: https://deepmind.google/models/gemini/

[6] (2026) The Arena AI website. [Online]. Available: https://arena.ai/leaderboard/vision

[7] (2026) The Google AI for Developers website. [Online]. Available: https://ai.google.dev/gemini-api/docs/models/gemini-3-flash-preview?hl=en

[8] W. Wang, J. Shen, F. Guo, M.M. Cheng and A. Borji, "Revisiting Video Saliency: A Large-scale Benchmark and a New Model," *Proceedings of the IEEE Conference on computer vision and pattern recognition*, pp. 4894-4903, 2018.

[9] P. K. Mital, T. J. Smith, R. L. Hill, and J. M. Henderson, "Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion," *Cognitive Computation*, New York: Springer Nature B.V., vol. 3, Iss. 1, Mar. 2011.

[10] A. Moskalenko, A. Bryncev, D. Vatolin, R. Timofte, G. Zhan, L. Yang, Y. Tang, Y. Liao, J. Lin, B. Huang, M. Moradi, M. Moradi, F. Rundo, C. Spampinato, A. Borji, S. Palazzo, Y. Zhu, Y. Nan, H. Duan, Y. Cao, Z. Jia, Q. Hu, X. Min, G. Zhai, H. Fang, R. Cong, X. Lu, X. Zhou, W. Zhang, C. Zhao, W. Mu, T. Deng, and H. R. Tavakoli, "AIM 2024 challenge on video saliency prediction: Methods and results," *European Conference on Computer Vision*, Cham: Springer Nature Switzerland, pp. 178-194, May 12, 2024.

[11] S. Mathe and C. Sminchisescu, "Actions in the eye: dynamic gaze datasets and learnt saliency models for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, Iss. 7, pp. 1408-1424, Jul. 01, 2015.

[12] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang, "Lost in the Middle: How Language Models Use Long Contexts," *Transactions of the Association for Computational Linguistics*, Vol. 12, p. 157, 2024.

[13] L. Gong, F. Bahrani, Y. Zhou, A. Banayeeanzade, J. Li, and E. Bıyık, "AutoFocus-IL: VLM-based Saliency Maps for Data-Efficient Visual Imitation Learning without Extra Human Annotations," *IEEE International Conference on Robotics and Automation (ICRA)*, 2026.

# Single-Image-Driven Automated 3D Avatar Generation and Real-Time Motion Retargeting Pipeline for Immersive Exhibition Interaction

G. Kim[1], Y. Shin[1], and S. Seo[1,*]
[1]*School of Art and Technology, Chung-Ang University, Anseong, 17546, South Korea*
* Contact: sanghyun@cau.ac.kr, phone +82-10 7273 0318

***Abstract*—Contemporary exhibition content is evolving beyond one-way information delivery toward immersive, participatory experiences in which visitors engage directly. Within virtual environments, personalized avatars play a crucial role in shaping users' sense of immersion and presence. However, existing approaches — such as manual customization or costly 3D scanning — increase visitor fatigue and face physical and temporal constraints that make them difficult to deploy in offline exhibition settings where real-time responsiveness is essential. To overcome these limitations, this study proposes an integrated pipeline that automatically generates a 3D avatar reflecting a visitor's physical appearance from a single input image in real time, and links it seamlessly to exhibition content. The proposed pipeline applies Stable Diffusion XL and IP-Adapter-FaceID to a visitor's facial image to extract facial features, then uses ControlNet to consistently generate a 2D avatar image in T-Pose. The 2D image is subsequently converted into a 3D mesh using the Trellis3D model, after which automatic rigging conforming to Unity Engine's humanoid specification is applied via the UniRig framework. At runtime, skeleton data captured from visitors via an Azure Kinect sensor is processed through inverse kinematics (IK) and a Kalman filter-like smoothing algorithm, enabling lag-free synchronization with the avatar's motion. Qualitative evaluation conducted with a participant group in an actual exhibition environment showed that the generated personalized avatars achieved high facial feature preservation scores, and that visitors experienced a sense of body ownership and agency through motion-based interaction. Furthermore, interaction with an avatar resembling oneself elicited a deep emotional bond, extending visitors' dwell time in the exhibition compared to those using preset avatars. To enable smooth deployment in offline exhibitions with unspecified large audiences in the future, model optimization and distributed processing architecture design will be required to reduce the current inference latency of approximately 60–80 seconds.***

## I. Introduction

Contemporary exhibition content is evolving beyond one-way information delivery toward immersive, participatory experiences in which visitors engage and interact directly. Amid this shift, XR technology has established itself as a key enabler for maximizing visitor immersion. Within virtual environments, the presence of an avatar representing the visitor exerts a significant influence on the user's sense of spatial presence and agency in virtual space. According to research in cognitive psychology and virtual environment interaction, the level of immersion in a virtual environment is largely determined by the degree of visual similarity between the user and their avatar, as well as the level of movement synchronization. In other words, using a personalized avatar that resembles the user has been shown to induce statistically significantly higher levels of immersion, emotional attachment, and willingness to interact compared to selecting a pre-made avatar [1].

However, the methods currently used to generate personalized avatars in commercial exhibition content and virtual platforms carry an inherent limitation: they disrupt the flow of the exhibition experience. The majority of platforms adopt a preset selection approach, in which users manually assemble a character by combining parts — such as hair, face shape, eyes, nose, and mouth — from a fixed set of pre-configured assets, choosing whichever combination they judge to most closely resemble their own appearance. This approach not only increases visitor fatigue during the creation process, but also makes it difficult to faithfully capture subtle individual facial features, which can produce a sense of alienation toward the virtual character or interfere with the user's psychological identification with the avatar. While high-quality personalized avatars can indeed be generated using 3D scanning booths or multi-camera photogrammetry, such methods require expensive hardware and demand considerable time for data processing and rendering. This presents a critical constraint for offline exhibition environments, where large numbers of visitors must experience the content in succession and real-time responsiveness is essential.

To overcome these physical and temporal constraints and deliver a personalized interactive experience to each visitor, this study proposes an automated pipeline that generates a 3D avatar reflecting the user's facial features in real time from a single image input, and immediately uploads it to the exhibition content for live operation. The pipeline proceeds through a series of steps: a single RGB image of the visitor's face is used to generate a 2D avatar image via a generative model, which is then converted into a 3D mesh, automatically rigged to conform to Unity Engine's humanoid specification, and uploaded to the Unity scene at runtime. The finalized avatar is synchronized with the visitor's real-time skeleton data via an Azure Kinect sensor, enabling interaction with the visitor within the exhibition space.

The pipeline is designed to allow users to effortlessly create a 3D avatar in their own likeness with minimal intervention, combining image generation, 3D mesh creation, and automatic rigging within a single unified system. This integrated pipeline is capable of preserving each individual user's identity with precision while maintaining a consistent visual style aligned with the artistic concept of a given exhibition, making it readily extensible to a wide range of future applications including metaverse platforms, digital twins, and interactive media art content.

## II. Methodology

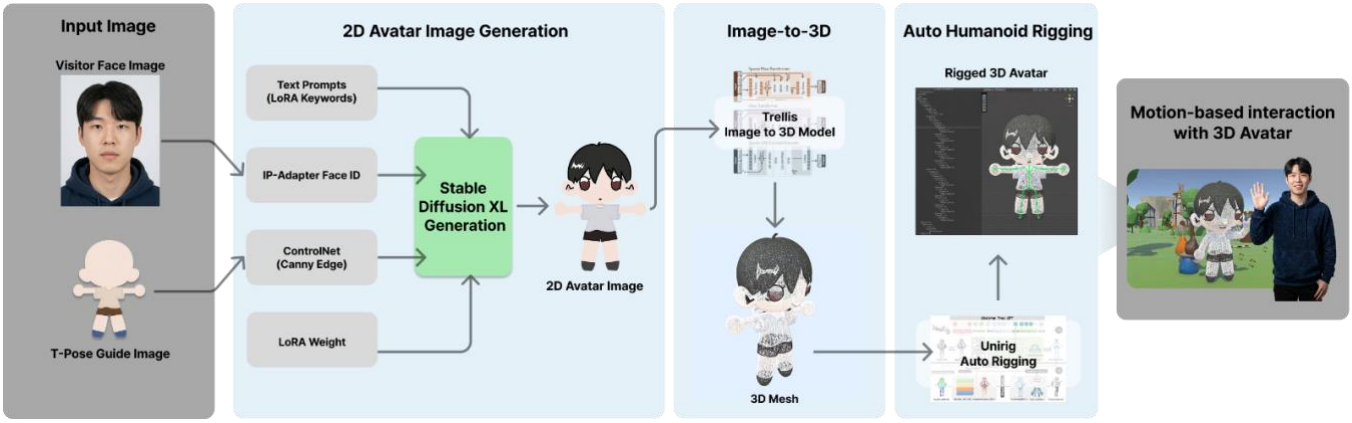



Fig. 1 Overall pipeline for motion-interactive 3D avatar generation. This pipeline transforms a single visitor face image into an interactive 3D avatar through three stages: 2D avatar generation using Stable Diffusion XL conditioned on IP-Adapter and ControlNet, 3D mesh reconstruction via Trellis, automatic humanoid rigging using Unirig.

In this study, the process of generating a 3D avatar that can be loaded into the Unity game engine and interact based on visitor input was designed across three processing modules. The overall pipeline is illustrated in Fig. 1, and the operating principles and detailed configuration of each module are as follows.

**1. 2D Avatar Image Generation Module**

A single facial image of the visitor is used to generate a 2D character image that aligns with the exhibition's artistic concept. Stable Diffusion XL was employed as the generative model [2]. To deliver a personalized experience, the visitor's facial features must be clearly reflected in the avatar. To this end, the IP-Adapter-FaceID architecture was employed [3]. Rather than using the conventional CLIP vision encoder, an InsightFace-based facial embedding model is used to extract only the visitor's facial features, which are then injected into the cross-attention blocks of SDXL [4]. This approach allows only facial identity — excluding background and clothing style — to be incorporated into the generation process, enabling avatar images to be produced while maintaining a consistent visual style suited to the content. The pose and body shape of the generated avatar image have a significant downstream impact on the subsequent 3D mesh conversion and rigging stages. To ensure that the 3D model is generated without overlapping geometry and that joints are accurately detected, a T-Pose or A-Pose configuration with clearly separated arms and legs must be enforced. In this pipeline, ControlNet was employed to control the generation of the 2D avatar image in T-Pose [5]. A control map was created by applying Canny edge extraction to a pre-prepared T-Pose reference image, which was then provided as an additional conditioning input to SDXL. This enables the consistent generation of 2D avatar images in which the face resembles the visitor while the pose conforms to T-Pose. To render a consistent visual style aligned with a specific exhibition art concept, LoRA (Low-Rank Adaptation) fine-tuning was applied to the Stable Diffusion XL base model (SDXL 1.0). The Network Rank (Dim) and Alpha values for model training were each set to 32, and training was performed using the AdamW optimizer

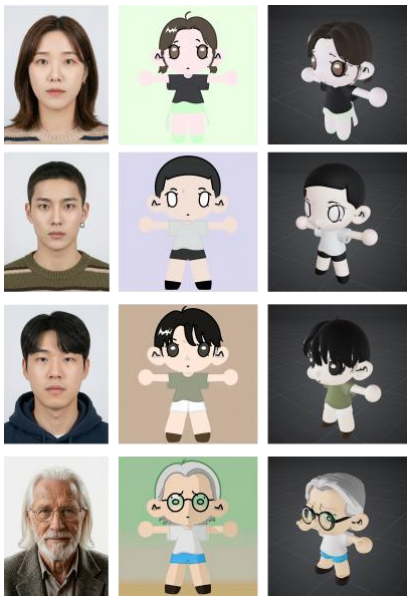


Fig. 2 Samples of generated 2D avatar image and 3D avatar based on visitor facial inputs.

with a learning rate of 3e-05 over a total of 50 epochs. The resolution of training images was fixed at 1024×1024, and FP16 mixed precision computation was applied for memory efficiency. The trained LoRA weights are integrated into the U-Net layers of the model. The resulting 2D avatar images produced by this module can be seen in Fig. 2.

**2. Image-to-3D Generation Module**

The 2D avatar image generated in the previous stage is fed as a reference image into an image-to-3D model and converted into a 3D mesh. In this pipeline, Trellis3D was used as the image-to-3D model [10]. The 3D results can be shown in Fig. 2.

**3. UniRig-Based Automatic Humanoid Rigging**

The generated mesh is merely a collection of vertices and faces and does not contain skeleton data necessary for applying animations. To address this, the automatic rigging

TABLE I
QUALITATIVE EVALUATION RESULTS COMPARING THE PRESET-BASED METHOD AND OUR PROPOSED APPROACH.

| Evaluation Cafegory | Survey Question | Scale | Preset based Mean | **Ours Mean** | **Improvement** |
|---|---|---|---|---|---|
| Overall Feature Preservation | How well does the generated avatar preserve your overall facial features? | 1 - 10 | 4.6 | 8.8 | +4.2 |
| Detailed Features (Eyes) | How accurately do the avatar's eyes reflect your unique characteristics? | 1 - 10 | 3.9 | 8.3 | +4.4 |
| Detailed Features (Facial Structure) | How well does the avatar's facial structure and shape represent your actual appearance? | 1 - 10 | 4.2 | 7.6 | +3.4 |
| Style and Impression | How naturally is your distinct impression integrated into the specific avatar style of the exhibition content? | 1 - 10 | 5.7 | 9.1 | +3.4 |
| Overall Satisfaction | How satisfied are you with the overall quality of the avatar generated from a single photograph? | 1 - 10 | 5.1 | 8.7 | +3.6 |
| **Total Average** | | | **4.7** | **8.5** | **+3.8** |

framework UniRig was incorporated into the pipeline [6]. UniRig is built on a deep learning architecture that predicts a detailed and topologically appropriate humanoid skeleton from the input 3D mesh. In this pipeline, a post-processing script was implemented to map the tree structure of the generated joint nodes to be compatible with Unity Engine's animation system, known as the "Humanoid Rig." As a result, the 3D mesh avatar processed through UniRig is exported with joint names and a topological structure compatible with Unity.

A preparatory process is required to reflect the visitor's movements in real time onto the 3D avatar loaded at runtime. The Microsoft Azure Kinect DK sensor serves as the interface for capturing the visitor's skeletal data [7]. A Unity script receives the 32 skeleton data points collected via the Kinect sensor and maps them onto the Unity humanoid rig tree previously constructed through UniRig. At this stage, rather than copying absolute positional coordinates, an IK processing technique is applied — using root joint translation and the relative rotation between each joint — in order to compensate for discrepancies between the avatar's skeletal proportions and those of the actual visitor. Additionally, a Kalman filter-like smoothing algorithm is applied to mitigate subtle trembling artifacts in the avatar caused by sensor recognition noise, yielding natural motion output. This enables an interaction in which, whenever a visitor waves an arm, bends their body, or performs any other action in front of the screen, the personalized avatar inside the screen mirrors that same action without delay.

## III. EXPERIMENT & RESULTS

To demonstrate the effectiveness of the motion interaction approach enabled by the single-image-based avatar generation automation pipeline, a qualitative performance evaluation was conducted with a participant group in an actual exhibition environment. Participants were asked to assign a score from 1 to 10, with higher scores indicating a stronger perception that the avatar's features were well preserved. The preset-based avatar received a score of 4.7. In contrast, the personalized avatar generated through the proposed pipeline achieved a high average feature reflection score of 8.5. Visitors responded enthusiastically to the fact that their distinctive eye shape, bone structure, and overall impression were captured and rendered in the avatar style of the content from just a single photograph. The survey questions and results can be found in TABLE I.

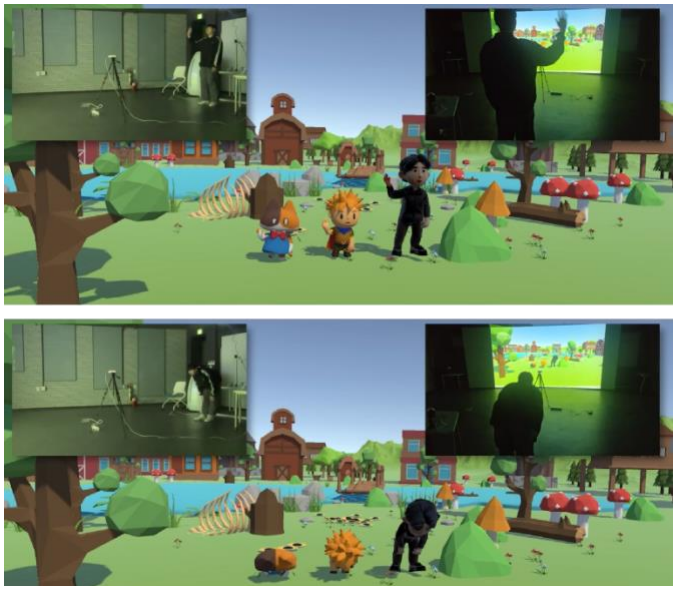


Fig. 3 Live exhibition test results of real-time 3D avatar motion retargeting. Top shows hand-waving; bottom shows bowing. Inset views present the physical user captured by the Azure Kinect.

After loading the avatar into the Unity environment and allowing participants to experience motion interaction via the Azure Kinect, qualitative interviews and behavioral observation analysis were conducted. This evaluation focused on the three primary experiential indicators of virtual

environments: presence, agency, and emotional connection. The majority of visitors reported experiencing a powerful sense of body ownership — akin to looking in a mirror — when the avatar on the screen instantly reproduced their movements. Fig. 3 illustrates this real-time motion synchronization, showing a physical posture being seamlessly and accurately mirrored by their personalized 3D avatar within the virtual exhibition environment. This experimentally reaffirmed existing psychological research findings suggesting that user agency is amplified when a visually self-resembling personalized avatar is combined with synchronized motion mapping technology [11]. Compared to controlling a generic avatar assembled from preset combinations, visitors smiled noticeably more often and attempted more exaggerated gestures when interacting with an avatar that resembled themselves. Synchronization with a digital object onto which personal identity was projected induced a deep emotional bond that went beyond mere entertainment, ultimately resulting in a meaningfully extended dwell time in front of specific exhibition content compared to the control group.

## IV. Conclusions

In this study, with the goal of devising an interaction method capable of elevating visitor immersion in an exhibition environment, a pipeline was successfully designed and validated that automatically generates a 3D avatar reflecting a visitor's unique facial features and clothing patterns from a single photograph, and drives it with real-time skeleton-based motion.

The proposed system combined IP-Adapter FaceID and ControlNet on top of SDXL to generate a T-Pose 2D avatar image with preserved subject features, which was then connected to an image-to-3D model and the UniRig framework. The resulting integrated data is instantiated in the Unity Engine runtime environment via an SCP shell script and the glTFast package, and is immediately synchronized with Azure Kinect depth tracking data to enable real-time interactive engagement.

This end-to-end integrated pipeline allows visitors to generate a personalized avatar that intuitively reflects their own appearance and individuality with a single photograph, free from the constraints of pre-configured preset templates. The underlying process holds potential for application beyond exhibition software, extending to metaverse ecosystems, digital twin construction, and customized interactive art.

However, in order to deploy this architecture in an offline exhibition environment where an unspecified number of visitors arrive in continuous succession, clear limitations must be overcome. The most significant issue is the inference time bottleneck that inevitably arises from chaining the sequential inference of a large-parameter diffusion model and a large transformer-based rigging model. The current latency of 60 to 80 seconds risks disrupting the flow from the moment a visitor takes their photograph to the moment they enter the scene. To address this, future research must explore the following optimization methodologies in depth.

First, on the generative model side, research should be conducted to shorten the 2D rendering time by introducing SD-Turbo series models or latent consistency model architectures capable of high-quality inference with only a small number of denoising steps, as alternatives to SDXL. Second, on the system architecture side, a cloud- or edge-computing-based pipelined distributed processing architecture should be designed that fully separates the 2D image computation node, the 3D rigging node, and the Unity rendering node, with load balancing applied across them.

Furthermore, if occlusion issues in simultaneous multi-visitor skeleton tracking using multiple sensors are resolved, and if an LLM-based voice recognition interface is integrated with avatar facial expression animation (Audio-to-Face) as a multi-modal extension, the proposed pipeline has the potential to advance into a foundational pipeline for next-generation immersive extended reality systems

## Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176)

## References

[1] Myeongul Jung, Sangyong Sim, Jejoong Kim, Kwanguk Kim, Impact of Personalized Avatars and Motion Synchrony on Embodiment and Users' Subjective Experience: Empirical Study, JMIR Serious Games, Volume 10, Issue 4, 2022

[2] D. Podell et al., "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis," arXiv preprint arXiv:2307.01952, 2023. [Online].

[3] H. Ye, J. Zhang, S. Liu, X. Han, and W. Yang, "IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models," arXiv preprint arXiv:2308.06721, 2023. [Online].

[4] J. Guo, J. Deng, X. An, and J. Yu, "InsightFace: 2D and 3D Face Analysis Project," GitHub repository, 2020. [Online].

[5] L. Zhang, A. Rao, and M. Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models," arXiv preprint arXiv:2302.05543, 2023. [Online].

[6] J.-P. Zhang, C.-F. Pu, M.-H. Guo, Y.-P. Cao, and S.-M. Hu, "One Model to Rig Them All: Diverse Skeleton Rigging with UniRig," ACM Transactions on Graphics (SIGGRAPH), vol. 44, no. 4, Art. 123, 2025. [Online].

[7] Z. Liu, "Azure Kinect Body Tracking SDK," Microsoft Research, 2020. [Online].

[8] H. Li et al., "On the Scalability of Diffusion-based Text-to-Image Generation," arXiv preprint arXiv:2404.02883, 2024. [Online].

[9] "Impact of Personalized Avatars and Motion Synchrony on Embodiment and Users' Subjective Experience: Empirical Study," JMIR Serious Games, vol. 10, no. 4, e40119, 2022. [Online].

[10] J. Xiang et al., "Structured 3D Latents for Scalable and Versatile 3D Generation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025. [Online].

[11] N. Yee and J. Bailenson, "The Proteus Effect: The Effect of Transformed Self-Representation on Behavior," Human Communication Research, vol. 33, no. 3, pp. 271-290, 2007. DOI: 10.1111/j.1468-2958.2007.0029

# Hierarchical Stress Detection from Autonomic Nervous System Signals in Immersive Virtual Reality Using Attention Mechanisms: A Transformer Approach

*Ahsan Aziz[1], Chomyong Kim[2], Yunyoung Nam[3]*

*[1]Department of ICT Convergence, Soonchunhyang University, Asan, 31538, Republic of Korea*

*[2] ICT Convergence Research Centre, Soonchunhyang University, Asan, South Korea*

*[3] Department of Computer Science and Engineering, Soonchunhyang University, Asan, South Korea*

**Contact: ynam@sch.ac.kr*

**Abstract:** The activation of the autonomic nervous system is of critical importance for the monitoring of emotional states of arousal, stress, and engagement in practical scenarios. Physiological signals provide a consistent, reliable, and interpretable measure of autonomic nervous system response, reflecting both sympathetic and parasympathetic activity. In this paper, the authors have proposed a deep learning framework using the transformer architecture for the binary classification of stress states using the publicly available WorkStress3D and WESAD datasets. The dataset was preprocessed using class filtering, normalization, and segmentation using the sliding window technique. In the proposed framework, the input data was transformed into a higher-dimensional feature space. A global average pooling technique was used for the reduction of sequence data into compact feature representations. An SVM classifier was used for the classification of the stress states. In the experimental analysis, the proposed transformer architecture was found to perform better than the conventional deep learning models. The proposed methodology was the use of the Hybrid CNN-Transformer model for the classification of stress states, which resulted in an accuracy of 0.9798, a precision rate of 0.9816, a specificity of 0.9897, an F1-score of 0.9719, and an AUC of 0.9983.

**Keywords:** Nervous system activation, Transformer, Physiological signal analysis

## I. Introduction

The Autonomic Nervous System (ANS) is essential in affective computing and medical studies due to emotional and cognitive states have a direct impact on physiological signals that the ANS controls. For instance, stress or anxiety may result in the heart rate and level of electrodermal activity go up. This makes ANS-related signals very useful for systems that recognize psychological states, keep an eye on mental health, and let people interact with conventional computers.

The WESAD (Wearable Stress and Affect Detection) dataset is a prevalent instrument to conduct research on distinguishing between emotions and stressful situations. It was acquired from fifteen participants utilizing wearable sensors, which included the RespiBAN chest device and the Empatica E4 wristband. The dataset contains a variety of physiological signals complicatedly related to an ANS activity, which include the following:

- Electrocardiogram (ECG)
- Electrodermal Activity (EDA)
- Respiration (RESP)
- Body temperature
- Accelerometer data.

The main contribution of our work is as follows:

Establish up an effective preprocessing pipeline with windowing and normalization, and resolved label anomalies by only retaining valid annotated segments. Established a binary Relaxed and Stress classifier the fact that corresponds with clinical use cases, and then enhanced training by using the Adam optimizer and modifying hyperparameters. While employing positional encoding to keep information about an arrangement of occurrences, it also used multi-head self-attention to recognize dependencies across signals. Emphasized the significance of signal fusion over single-modality approaches and recommended a design that could expand to incorporate wrist signals (E4 device). Demonstrated that deep learning can be employed to detect stress in wearable technology. It additionally offered a structured framework that could potentially be simple to implement with other physiological datasets. Set a baseline for subsequent advanced models like MaxViT, TimesNet, and InceptionTime. A CNN-Transformer architecture that is a mix of the two and can capture local and global dependencies. Associated the opening between physiological data analysis and AI-driven classification. Shown wide evaluation using Accuracy, Precision, Recall, Specificity, and F1 score.

## II. Related Work

The WorkStress3D dataset [1] has only recently grown a significant component of research on stress analysis because it gives investigators a comprehensive framework for understanding stress at work. The dataset incorporates physiological signals, facial expressions, and speech data, allowing researchers to look at stress from both biological and behavioral perspectives. view. Traditional datasets only look at one type of data, but this one looks at electrodermal activity, blood volume pulse, and skin temperature in addition to visual and audio cues. This gives an improved understanding of how stress affects people. One of the best everything about the dataset is that it is ecologically valid, which means that the data were collected in real workplaces as opposed to controlled labs. The above technique lets you see stress that happens naturally, which makes the dataset more like what transpires in real life. Another significant aspect [2] raised in the literature is that the dataset can incorporate data gathered from multiple sources, which helps us learn more about how to detect stress. Researchers can look for links between physiological and behavioral signs of stress by combining biosignals, facial expressions, and speech. This design improves classification performance and makes it feasible to create more advanced affective computing platforms. The dataset is additionally provided with detailed metadata and notes to assist with feature extraction, pattern recognition, and training machine learning and deep learning models to classify stress, using an experience sampling method, where participants fill out questionnaires in real time, and their physiological data is constantly being tracked. Wearable devices like the Empatica E4, along with regular questionnaires and naturally occurring recordings of facial and vocal expressions, were used for collecting data. This methodology consents for capturing temporal variations in stress throughout daily activities without artificially persuading stress, thus preserving authenticity. As a result, WorkStress3D serves as a appreciated benchmark for evaluating stress detection algorithms and progressing research in workplace mental health and human-centered AI systems.

## III. Proposed Methodology

### A. Dataset

The WorkStress3D dataset (hosted on Mendeley Data) is a dataset designed to investigate acute mental stress under realistic work-pressure conditions. The dataset's main goal is to record individual physiological, behavioral, and contextual reactions when they complete challenging cognitive tasks under time constraints and performance evaluation. WorkStress3D incorporates cognitive workload, emotional arousal, and performance feedback in an effort to replicate real-world work pressure, in contrast with strictly synthetic stress paradigms. Twenty participants' recordings have been included in the dataset. Each participant underwent multiple task sessions (usually 41 per subject) covering baseline, task engagement, and anxious conditions. Researchers can examine interpersonal variability, inter-subject differences, and stress generalization models through this repeated-session design.

Both physiological signals and contextual task-related elements are incorporated in the data collection protocol. Time pressure, performance scoring, and potentially evaluative feedback were in the controlled stressors that were introduced into the structured experimental tasks that participants completed. The dataset comprises video-related elements that match the task environment or experimental stimuli in addition to physiological recordings. These videos typically depict work-like situations meant to cause emotional and cognitive strain. The multifaceted nature of stress assessment appears in WorkStress3D's "3D," which combines behavioral indicators, task context, and physiologic responses. Timestamp markers are used to coordinate task events with physiological recordings, allowing accurate initial and stress interval segmentation.

Physiological monitoring with the Empatica E4 wristband, a medical-grade wearable that is frequently used in affective computing and stress research, is a key part of the dataset. various CSV files are recorded for every session (file1 to file41 inside each user folder). The electrodermal activity sampled at low frequency (usually 4 Hz) in EDA.csv evaluates changes in skin conductance linked to sympathetic nervous system activation, one of the most powerful biomarkers of stress. Heart rate and heart rate variability (HRV) can be derived from the blood volume pulse data (about 64 Hz) stored in BVP.csv. While IBI.csv offers the inter-beat interval information needed for advanced HRV analysis (such as RMSSD and SDNN), HR.csv contains processed heart rate values in beats per minute. TEMP.csv records peripheral skin temperature, which can decrease under stress due to vasoconstriction. ACC.csv contains tri-axial accelerometer data (X, Y, Z) used mainly for movement detection and artifact removal. Finally, tags.csv contains timestamp markers corresponding to experimental events such as baseline start, stress block onset, or task transitions. These tags

are crucial for labeling segments as “Non-stress” or “Stress” in machine learning pipelines.

The dataset contains Empatica signals together with structured metadata that documents the experimental setup and anonymized participant details and any questionnaire assessments which were collected during or after the experiment. The questionnaires assess perceived stress levels and workload scores and gather personal feedback on how difficult the tasks were. The metadata allows researchers to analyze the relationship between actual physical body responses and how people experience stress. The study implemented a 41-session design which let participants complete different task sessions through their entire study. The first study delivers all the needed data in one zip file which contains organized data for every subject and every task.

WorkStress3D serves as an important research tool for researchers who examine stress detection methods and affective computing systems and human performance assessment and wearable artificial intelligence technologies. High-resolution cardiovascular data, BVP and IBI, together with autonomic arousal indicators EDA and peripheral thermoregulation TEMP and motion artifacts, ACC and event tags and contextual task/video stimuli create a comprehensive system which enables deep learning. Your hybrid transformer approach requires EDA.csv and BVP.csv and IBI.csv and tags.csv as its most essential files because these files show autonomic nervous system behavior and deliver the actual stress classification boundaries. The dataset’s structured session-based design enables effective sequence modeling and cross-session validation and development of personalized stress prediction systems.

WESAD was collected under measured laboratory conditions using both wearable and clinical-grade devices, providing high-quality physiological signals such as ECG, PPG, EDA, EMG, respiration, temperature, and accelerometer data. It includes multiple subjects exposed to diverse affective states such as neutral, stress, and amusement, making it highly valuable for binary and multi-class classification tasks. The dataset comprises 15 participants (after excluding 2 due to sensor malfunction). Mean age ~ 27.5 ± 2.4 years; 12 male, 3 female, and exclusion criteria are pregnancy, heavy smoking, mental/cardiovascular disorders, etc. Devices and sensors used are Chest-worn: RespiBAN Professional sensors: ECG, EDA, EMG, RESP (respiration), TEMP (body temperature), 3-axis ACC (accelerometry), all sampled at 700 Hz. Wrist-worn: Empatica E4 sensors: BVP (blood volume pulse / PPG) at 64 Hz, EDA at 4 Hz, TEMP at 4 Hz, ACC at 32 Hz. with sensor placement at RespiBAN: chest; analog ports for ECG/EDA/EMG/TEMP; electrodes and sensors on chest, abdomen (for EDA), sternum (TEMP), upper trapezius (EMG), etc. Empatica E4: non-dominant wrist. The duration of the experiment is Baseline ~20 minutes; Amusement ~392 seconds (~6.5 minutes) (with interleaving neutral parts between clips), Stress condition ~10 minutes (speech + counting task). Meditation/rest periods are also included. Total experiment per participant ~2 hours. Sample rates for Chest device (RespiBAN): 700 Hz for all chest signals and Wrist device (Empatica E4): BVP 64 Hz; EDA & TEMP 4 Hz; ACC 32 Hz. The signals collected are ECG, EDA, EMG, Respiration (RESP), Body Temperature, Accelerometry (3-axis) from chest; BVP, EDA, Temperature, ACC from the wrist.

## B. Dataset Organization & Signal Handling

The proposed framework uses the WorkStress3D dataset, which contains physiological recordings collected with the Empatica E4 wearable device. In the current implementation, the dataset is organized in a flat directory structure where each session file follows a standardized naming format such as UserX_fileY__EDA.csv, BVP.csv, TEMP.csv, and related signals. The implemented data loader automatically scans the root directory, groups files by user-session combination, and pairs relevant physiological signals (EDA and BVP) for further processing. MacOS metadata files and incomplete entries are programmatically excluded to ensure data integrity. This automated grouping mechanism enables scalable processing of large session-based physiological datasets without manual intervention. Each Empatica CSV file follows a standardized format in which the first row contains the start timestamp, the second row specifies the sampling frequency, and the remaining rows contain the actual signal values. The implemented loader correctly removes the metadata rows and extracts only the physiological signal values. EDA (Electrodermal Activity) and BVP (Blood Volume Pulse) are selected as primary stress-related biomarkers due to their strong relationship with autonomic nervous system activation. After extraction, both signals are temporally aligned by truncating them to the minimum shared length within each session.

## C. Signal Preprocessing and Normalization

To ensure robust learning and inter-session generalization, the physiological signals are normalized independently within each session using

z-score normalization. This step removes subject-specific amplitude bias and reduces the impact of baseline physiological variability. Since EDA and BVP operate at different sampling frequencies in raw Empatica recordings, alignment is achieved by length truncation in the current implementation; however, the structure allows future integration of advanced resampling techniques if needed. Following preprocessing, the signals are concatenated channel-wise, forming a multivariate time-series representation:

$$X \in R^{T \times C}$$

Where $T$ represents temporal samples and $C = 2$ corresponds to the physiological channels (EDA and BVP). This multichannel configuration enables the model to learn cross-signal dependencies rather than treating each physiological source independently.

## D. Window-Based Segmentation Strategy

To transform continuous recordings into machine learning-ready samples, a sliding window segmentation approach is implemented. Fixed-length windows (e.g., 256 samples) with configurable overlap (step size = 128) are generated. Each window is assigned a class label based on the midpoint of the window. In the current version, a provisional binary labeling strategy is used, where we have each session assigned to the non-stress class and the stress class. This structure ensures balanced segmentation while maintaining temporal continuity. Windowing serves multiple purposes, such as increasing dataset size for deep learning training, preserving temporal structure within each segment, enabling localized stress pattern learning, and reducing memory load during training.

The final dataset tensor has the shape:

$$N, Window_{size}, channels$$

which is directly compatible with the transformer-based sequence models.

## E. Transformer-Based Temporal Modeling

The design recommended a Time-Series Transformer encoder that is made for classifying multivariate physiological patterning. First, the input features are projected linearly into a fixed embedding dimension of 64. Then, the projected sequences are processed by stacked Transformer encoder layers that use multi-head self-attention mechanisms. The self-attention module lets the model pick up on long-term temporal dependencies and interactions among signals, which is very important for modeling physiological stress because autonomic responses may show delayed or dispersed temporal patterns.

The transformer processes the whole sequence at once, thereby rendering training faster and helping it learn global contextual dependencies better than recurrent architectures like LSTM and GRU. Global average pooling is used across the time dimension to get a fixed-length representation. This is then sent to a completely interconnected classifier that produces binary outputs (Relaxed and Stressed). The model is trained using cross-entropy loss and optimized via the Adam optimizer.

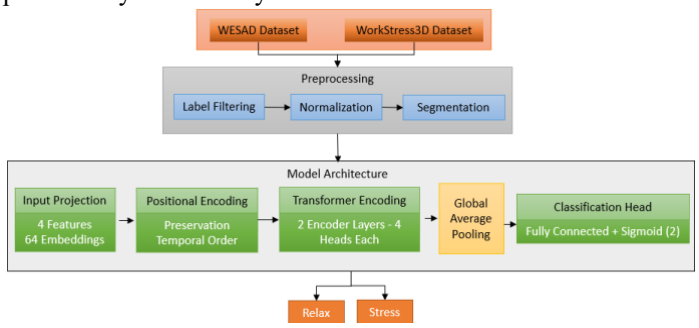



**Figure 1: Proposed Flow Chart Based on Transformer**

Also, the core network had two stacked Transformer encoder layers: The first component is the multi-head self-attention mechanism with four heads that detects dependencies among features as well as among time periods. The second part is the feed-forward network with 128 hidden units. In order to make the model more general, Dropout and layer normalization are used. A global average pooling layer then integrates the outputs at the sequence level into a fixed-size representation. The classification head had a fully connected layer that associated the pooled vector to the number of output classes. For binary classification, these classes are generally thought of as two. The proposed architecture is illustrated in Figure 1.

## F. Hybrid CNN-Transformer

To generate a hybrid CNN-Transformer architecture that captures both local and global dependencies, start by integrating a convolutional front-end for local pattern extraction with a transformer for global\long-range dependencies, attention pooling, and a classification head. Where numerous stacked 1D convolutions, in addition to ReLU and max-pooling x 3 layers, are used to generate local features like ECG wave shapes and small transients and reduce the length of the sequence. Then, positional encoding is added to the outputs generated by the convolution, and multi-head attention with normalization and FFN blocks will capture long-range dependencies across time and features for transformer modules. Over time, Self-Attention Pooling will learn the procedure for the pool, which focuses on segments that provide standard information (like peaks and bursts). Extra Transformer/attention blocks, which are a second stage of attention that assists in enhancing features before classification.

## IV. Experimental Results

This study assessed the performance of a deep transformer architecture for the WESAD and WorkStress3D datasets.

**Table 1 Results using transformer-based binary classification**

| Model | Loss Function | Epochs | Batch-Size | Learning Rate | Precision Rate | Specificity Rate | F1-Score | AUC | Accuracy |
|---|---|---|---|---|---|---|---|---|---|
| **WESAD** | | | | | | | | | |
| **Transformer** | Cross-Entropy | 2 | 4 | 1e-4 | 0.8128 | 0.9502 | 0.5175 | 0.7966 | 0.7431 |
| | Cross-Entropy | 5 | 64 | 1e-4 | 0.8749 | 0.9706 | 0.5113 | 0.8154 | 0.7494 |
| | Cross-Entropy | 20 | 64 | 1e-4 | 0.8892 | 0.9744 | 0.5133 | 0.7837 | 0.7517 |
| **Hybrid-CNN-Transformer** | Cross-Entropy | 20 | 64 | 1e-4 | 0.9816 | 0.9897 | 0.9719 | 0.9983 | 0.9798 |
| **WorkStress3D** | | | | | | | | | |
| **Transformer** | Cross-Entropy | 2 | 4 | 1e-4 | 0.7782 | 0.7908 | 0.7309 | 0.831 | 0.7625 |
| | Cross-Entropy | 5 | 64 | 1e-4 | 0.7087 | 0.7562 | 0.7352 | 0.935 | 0.7603 |
| | Cross-Entropy | 20 | 64 | 1e-4 | 0.7714 | 0.7878 | 0.7536 | 0.876 | 0.7747 |
| **Hybrid-CNN-Transformer** | Cross-Entropy | 20 | 64 | 1e-4 | 0.9079 | 0.9488 | 0.8131 | 0.9346 | 0.8656 |

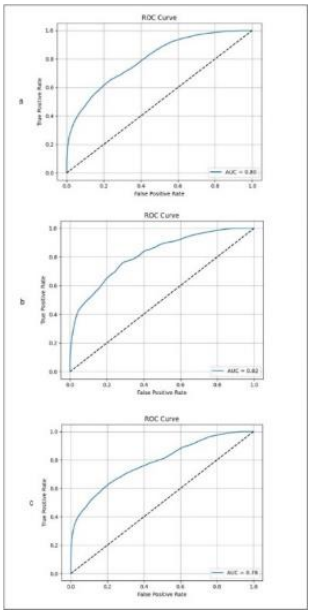



**Figure 2: AUC curves on the Transformer-based model for binary classification on the WESAD Dataset**

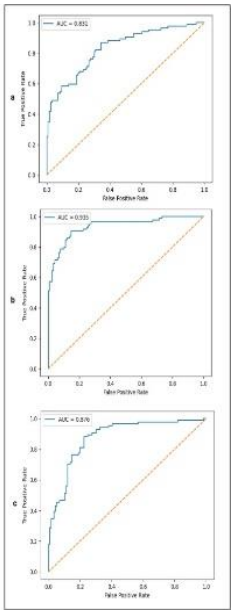


**Figure 3 AUC curves on the Transformer-based model for binary classification on the WorkStress3D Dataset**

In each figure 2 and 3, (a) AUC is represented as 2 epochs and 4 batch sizes, (b) as 5 epochs and 64 batch sizes, and (c) as 20 epochs and 64 batch sizes. In all aspects, the transformer-based method consistently demonstrated high specificity and precision, which demonstrates that it is reliable for finding Stress states. The model reached its highest accuracy (0.8892) and specificity (97.44%) after 20 epochs with the same batch size. The accuracy improved a little, to 75.17%. Our proposed method was a hybrid CNN-Transformer architecture that captures both local and global dependencies. An accuracy of 0.9798, a Precision Rate of 0.9816, a Specificity Rate of 0.9897, an F1-score of 0.9719, and an AUC of 0.9983. Figures 4 and 5 illustrate the confusion matrix and AUC curve for binary classification implementing our hybrid methodology. The high performance on WESAD can be attributed to the availability of high-quality synchronized physiological signals, official annotations, and robust preprocessing procedures. Furthermore, transformer-based temporal modeling effectively captured stress-related physiological patterns, and we used an 80:20 train/test split.

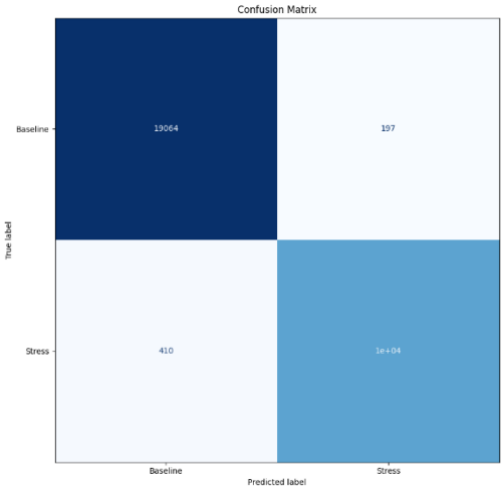



**Figure 4: Confusion Matrix on Hybrid CNN-Transformer for Binary Classification on WESAD Dataset**

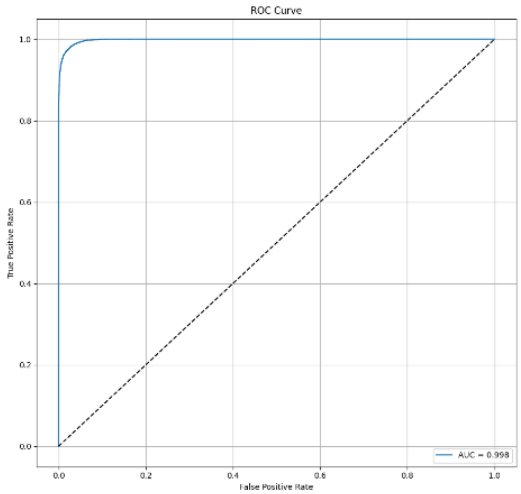



**Figure 5: AUC-ROC Curve on Hybrid CNN-Transformer for Binary Classification**

## V. Discussion

The transformer makes an output for each time step in the window. Global average pooling incorporates all of these outputs into one feature vector that is always the same length, making it suitable for classification.

subsequently achieving this, the features go through a fully connected layer, and sigmoid activation is used to sort them into two groups. We then indicated a hybrid CNN-Transformer model that can learn local and global dependencies. The model has a convolutional front-end, a transformer module, attention pooling, and a classification head. We utilized stacked 1D convolutional layers, ReLU, and max pooling x 3 layers to learn local features like ECG wave shapes and small transients, and to reduce the sequence length. Following that, positional encoding has been added to the output of the convolutional neural network. The transformer module can learn long-range dependencies through the use of the multi-head attention module, along with normalization and the FFN module. Self-Attention Pooling can also learn to pool over time, concentrating on important parts of the ECG signals, including the peaks and bursts. There are more transformer/attention blocks in the second stage of the attention mechanism. These blocks learn the features before the classification head. The fully connected layer converts the data, and subsequently, the SVM classifier provides the prediction task. The proposed hybrid model possesses an accuracy of 0.9798, a precision of 0.9816, a specificity of 0.9897, an F1-score of 0.9719, as well as an AUC of 0.9983.

## VI. Conclusion & Future Work

This work introduces a comprehensive framework for the binary classification of the autonomic nervous system (ANS) employing physiological data from the WorkStress3D and WESAD datasets. Signals like EDA, ECG, respiration, and temperature are incorporated into one input and processed by a Transformer-based model to find patterns over time that can be used to tell the difference between a stressed state and a relaxed state. A preprocessing pipeline with windowing, normalization, and label filtering makes sure that the data is accurate. Fusion, on the contrary aspect, works more effectively than single-signal methods. A hybrid architecture that combines convolutional layers with Transformer-based attention is additionally suggested to capture both local and global dependencies in physiological signals. The model has been evaluated on a subject-independent basis using standard metrics. It does an adequate job of identifying stress, but it needs to get better at classifying relaxed situations. The framework is an excellent place to begin for real-time stress monitoring with wearable devices that can be expanded up to more advanced models like MaxViT and TimesNet. We aim to create a complete controlled stress dataset, which we will gather through an immersive virtual reality (VR) environment using Empatica E4 and Meta Quest Pro equipment. The proposed dataset will be created through virtual reality scenarios, which researchers designed to produce authentic emotional and bodily responses, whereas existing datasets use standard laboratory stress tests that include arithmetic tasks and interviews. Participants will undergo structured enrollment procedures with ethical approval and informed consent. The Empatica E4 device will record synchronized physiological signals, which include electrodermal activity (EDA), blood volume pulse (BVP), inter-beat interval (IBI), and skin temperature. The Meta Quest Pro system will capture facial dynamics and behavioral cues through its built-in sensing functions. The system will achieve exact time synchronization to match the start of VR stimuli with the collection of various physiological data.

The VR stimuli will be categorized into two primary classes: stress and relaxation, with further subcategories designed to induce differentiated autonomic responses. Stress conditions will include high-speed motion environments, social-evaluative stress scenarios, competitive task simulations, and virtual danger-based experiences (e.g., height exposure or threat simulation). In contrast, relaxation conditions will incorporate soft visual environments such as ocean waves, forest landscapes, guided breathing scenes, and slow-motion ambient environments. This controlled categorization allows for hierarchical labeling (binary and multi-class), enabling both coarse-grained and fine-grained stress modeling. Additionally, stimulus intensity levels will be systematically varied to study graded stress responses rather than simple binary classification, which is a limitation in many existing datasets.

The uniqueness of the proposed dataset lies in its immersive, and ecologically valid design. Unlike datasets such as WESAD or other lab-based recordings that primarily rely on static tasks and limited environmental immersion, this dataset integrates wearable biosignals, facial behavioral cues, and immersive VR stimuli within a synchronized framework. The inclusion of dynamic VR-induced stressors creates realistic emotional engagement, potentially yielding richer autonomic patterns and stronger generalization capability for deep learning models. Furthermore, the combination of physiological signals with facial response data enables cross-modal analysis, making it suitable for hybrid transformer-based architectures and fusion research. To our knowledge, few publicly available datasets combine immersive VR, wearable autonomic signals, and facial dynamics within a unified stress-recognition framework, positioning the proposed

dataset as a one-of-a-kind resource for next-generation affective computing and wearable mental health monitoring systems. Table 2 shows the environment of the VR that we will be selecting, and in the figure 6, the future proposed flow chart of data collection is shown.

Table 2: VR Environment for relaxed and stressed

| | Stress | Relax |
|---|---|---|
| 1 | **Heights & Vertigo** [Standing on skyscraper edges, Glass bridges, Rope bridges over valleys, Elevator malfunction scenarios] | **Natural Landscapes** [Forest walks, Mountains & valleys, Lakes and rivers, Snowfall scenes] |
| 2 | **Speed & Motion** [Roller coasters, VR racing (cars, bikes, jets), Free-fall simulations] | **Ocean & Water Scenes** [Beaches with waves, Underwater coral reefs, Slow River floating] |
| 3 | **Social Stress** [Public speaking to a large virtual audience, Job interview simulations, Classroom oral exam scenarios] | **Guided Meditation & Mindfulness** [Breathing exercises, Body scan meditation, Floating in space or clouds] |
| 4 | **Threat & Danger** [Horror environments (dark rooms, sudden noises), Being chased, Natural disasters (earthquake, fire evacuation)] | **Soft Visual Motion** [ Slow clouds moving, Aurora borealis, Fireflies at night] |
| 5 | **Time Pressure & Task Load [Games]** | **Animal & Nature Interaction** [Dolphin's swimming<br><br>Birds flying, Pet interaction (cats, dogs)] |

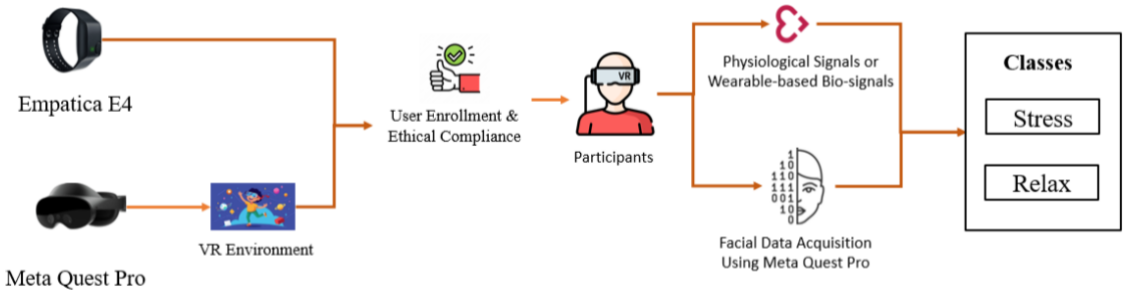



Figure 6: Future proposed flow chart of data collection

## Acknowledgment:

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176).

## VII. References

[1] Dogan G, Akbulut FP, Catal C. Biosignals, facial expressions, and speech as measures of workplace stress: Workstress3d dataset. Data Brief. 2024 Mar 8;54:110303. doi: 10.1016/j.dib.2024.110303. PMID: 38559821; PMCID: PMC10981000.

[2] Dogan, Gulin, Fatma Patlar Akbulut, and Cagatay Catal. "Biosignals, facial expressions, and speech as measures of workplace stress: Workstress3d dataset." Data in Brief 54 (2024): 110303.

[3] Hnoohom, Mekruksavanich & Jitpattanakul — "Physical Activity Recognition Based on Deep Learning Using Photoplethysmography and Wearable Inertial Sensors" (Electronics, 2023)

[4] Mekruksavanich, Hnoohom et al. — "Classification of Physical Exercise Activity from ECG, PPG and IMU Sensors using Deep Residual Network" (RI2C, 2022)

[5] Okamoto et al. — "Estimating Upper-Arm Muscle Activities and sEMG from Wrist PPG" (Sensors / ACM / MDPI, 2021–2023 work)

[6] Rashid et al. — "Feature-Augmented Hybrid CNN for Stress Recognition Using Wrist-based Photoplethysmography Sensor" (arXiv / preprint, 2021)

[7] Elzeiny & Qaraqe — "Stress Classification Using Photoplethysmogram-Based Spatial and Frequency Domain Images" (Sensors, 2020)

[8] Mohino-Herranz et al. — "Activity Recognition Using Wearable Physiological Signals" (Sensors, 2019) — (ECG / EDA / TEB)

[9] Shu, Lin, Yang Yu, Wenzhuo Chen, Haoqiang

Hua, Qin Li, Jianxiu Jin, and Xiangmin Xu. "Wearable emotion recognition using heart rate data from a smart bracelet." *Sensors* 20, no. 3 (2020): 718.

[10] Li, Xue, Chiaki Ono, Noriko Warita, Tomoka Shoji, Takashi Nakagawa, Hitomi Usukura, Zhiqian Yu et al. "Heart rate information-based machine learning prediction of emotions among pregnant women." *Frontiers in psychiatry* 12 (2022): 799029.

[11] Li, Xue, Chiaki Ono, Noriko Warita, Tomoka Shoji, Takashi Nakagawa, Hitomi Usukura, Zhiqian Yu et al. "Comprehensive evaluation of machine learning algorithms for predicting sleep–wake conditions and differentiating between the wake conditions before and after sleep during pregnancy based on heart rate variability." *Frontiers in Psychiatry* 14 (2023): 1104222.

[12] Debnath, Uday, and Sungho Kim. "A comprehensive review of heart rate measurement using remote photoplethysmography and deep learning." *BioMedical Engineering OnLine* 24, no. 1 (2025): 73.

[13] Salinas, Ruben F. *The Virtual Overview Effect: Exploring the Potential of Virtual Reality Simulations of Earth Gazing to Induce Awe and Measurable Changes in Heart Rate Variability*. Sofia University, 2023.

[14] Makowski, Dominique, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and SH Annabel Chen. "NeuroKit2: A Python toolbox for neurophysiological signal processing." *Behavior research methods* 53, no. 4 (2021): 1689-1696.

[15] Atiea, M. A., & Adel, M. (2022). Transformer-based neural network for electrocardiogram classification. *International Journal of Advanced Computer Science and Applications*, *13*(11).

[16] Darbhasayanam, S. H., Mohammad, H. A. A., & Komalla, A. R. (2024, May). Photoplethysmogram (PPG)-Based Blood Pressure Estimation using Vision Transformer Networks: A Deep Learning Approach. In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (pp. 1-6). IEEE.

[17] Ali, K., & Hughes, C. E. (2023). A unified transformer-based network for multimodal emotion recognition. *arXiv preprint arXiv:2308.14160*.

[18] Liu, Z. D., Zhou, B., Liu, J. K., Zhao, H., Li, Y., & Miao, F. (2024, July). A CNN and Transformer Hybrid Network for Multi-Class Arrhythmia Detection from Photoplethysmography. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1-5). IEEE.

[19] Ding, C., & Wu, C. (2024). Self-Supervised Learning for Biomedical Signal Processing: A Systematic Review on ECG and PPG Signals. *medRxiv*, 2024-09.

[20] Bastakoti, N., Poudel, S., Lamsal, S., Sharma, Y., & Jaiswal, A. Heart Rate Estimation from PPG and Accelerometer via CNN-Transformer Fusion: HeartFusion.

[21] Le, T. D. (2023). GRN-transformer: Enhancing motion artifact detection in PICU photoplethysmogram signals. *arXiv preprint arXiv:2308.03722*.

[22] Ahsan Aziz, Chomyong Kim, Yunyoung Nam “Deep Transformer Architecture for Physiological Signal Analysis in Autonomic Nervous System State Classification of Active and Non-Active”

# The Effects of Family Relationships on Social Relationship Satisfaction in Children and Adolescents with Developmental Disabilities

Sanggyeong Lee [1], Joungmin Kim[2], and Yaena Ha[3,*]
[1]*ICT Convergence, Graduate School, Soonchunhyang University, Republic of Korea*
[2]*Hyangseol College of Liberal Arts, Soonchunhyang University, Republic of Korea*
[3]*Emotional and Intelligent Child Care System Convergence Research Center, Soonchunhyang University, Republic of Korea*
*Contact: otha1003a@sch.ac.kr

*Abstract*— **This study examined how family relationships influence social relationship satisfaction among students with developmental disabilities and compared these effects by developmental stage. Using raw data from the 5th wave (2022) of the Disability and Life Dynamics Panel Survey by KODDI, 211 students with developmental disabilities were analyzed, including 148 children and 63 adolescents. Descriptive statistics, independent samples t-tests, and multiple regression analyses controlling for general characteristics were conducted using R 4.5.1. The child group scored significantly higher than the adolescent group in "Protection and Support" and "Clarity of Rules and Values," indicating weaker perceptions of family closeness and shared norms in adolescence. Family communication positively affected social relationship satisfaction in both groups, but its specific effects differed by developmental stage: honesty and decision-making participation were important in childhood, whereas communication and trust were significant in adolescence. Family problem-solving capability showed a negative effect in adolescents. These findings suggest a shift from communication-centered support in childhood to autonomy-respecting support in adolescence.**

## I. INTRODUCTION

Developmental disabilities (DD) entail continuous difficulties in overall daily living, including learning, communication, and social adaptation [1], [2]. According to the 2025 Special Education Statistics, students with DD constitute 70.50% of special education recipients in South Korea, reflecting a steady increase from 68.30% in 2021 [3].

For children and adolescents with DD, the family serves as the primary environment and a core support system carrying out the most intimate interactions throughout the life cycle [4], [5]. Due to cognitive and communicative challenges, individuals with DD exhibit a high dependence on their families, leading to lifelong parental caregiving burdens [6]. According to Bronfenbrenner's ecological systems theory, the family is the initial microsystem, serving as the primary environment where foundational competencies for social adaptation are acquired [7]. Hence, positive family relationships can function as an emotional buffer against negative experiences and support successful social integration [7].

Healthy family relationships provide a critical cornerstone for the successful social adaptation of students with disabilities [8], [9]. Yoo defined healthy family relationships as a system that promotes the development of individual members, shares a common value system, and achieves smooth interactions with wider social systems [10]. This framework consists of four sub-domains: family cohesion, family communication, shared family value system, and family problem-solving capability [11], [12]. Recent literature reports that higher levels of family functioning are associated with enhanced social adaptation [13]. Conversely, the difficulties in social relationships experienced by students with DD often trigger peer alienation and school maladjustment, deteriorating overall life satisfaction [14]. If parents maintain authoritarian control or an overprotective attitude as their children grow, it deprives the children of opportunities for self-determination [15]. Also, caregiving conflicts within the family can cast a negative shadow over general domestic and social life [16]. Therefore, improving social relationship satisfaction among students with DD requires closer attention to the characteristics of their family relationships.

Several limitations exist in previous studies addressing the link between the family relationships of students with DD and their social adaptation. First, a vast majority of research focuses on the parental perspective, leaving a relative scarcity of studies directly analyzing family relationships as perceived by the individuals with disabilities themselves. Second, while existing studies verified the impacts of family relationships on school adjustment [14], [17] or life satisfaction [18], empirical verification targeting social relationship satisfaction as a direct dependent variable remains insufficient. Third, research comparing developmental stage differences between childhood (elementary school) and adolescence (middle/high school) among students with DD is highly limited [17]. While dense protection acts as a core function during childhood, entering adolescence prompts an increase in autonomy demands, requiring a qualitative transformation in family interaction patterns [17].

The purpose of this study is to elucidate the effects of family relationships on social relationship satisfaction among school-aged students with DD using data from the 5th wave (2022) of the Disability and Life Dynamics Panel Survey by KODDI, and to comparatively analyze the differences in these effects between childhood (elementary) and adolescence (middle/high). The specific research questions are as follows:

First, is there a significant difference in family relationships between the child and adolescent groups of students with DD? Second, how do family relationships affect the social relationship satisfaction of these students, and does the magnitude of this impact differ between childhood and adolescence?

## II. Methods

### A. Data

This study utilized data from the 5th wave (2022) of the Disability and Life Dynamics Panel Survey conducted by KODDI [19]. This survey is a nationally approved statistical tool established to track the life trajectories of individuals registered with the Ministry of Health and Welfare after the onset of a disability. It comprehensively investigates daily life, income levels, health status, welfare needs, and social participation to provide foundational data for subsequent policy formulation [19].

### B. Participants

The final analysis sample was selected from the panel survey's 4,904 total respondents, filtering for 211 students with developmental disabilities (ID and ASD) currently enrolled in elementary, middle, or high schools (Fig. 1). In this study, students in elementary school were classified as children, while those in middle and high schools were classified as adolescents. The sample comprised 148 children (70.10%) and 63 adolescents (29.90%).

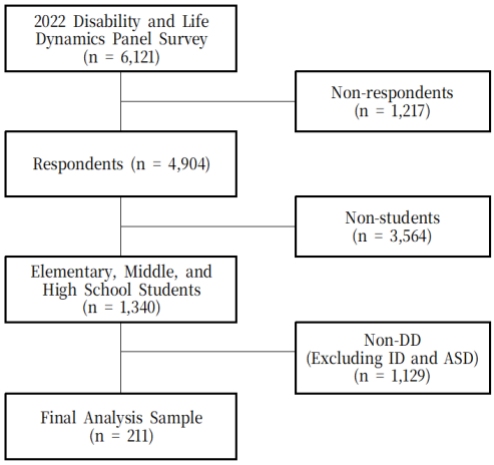



Fig. 1 Flowchart of the participant selection process.

### C. Variables

- Independent Variable: Family Relationships

Family relationships were measured using the Korean Family Strength Scale adopted by the panel survey [11]. The scale consists of 4 sub-domains—family cohesion, family communication, shared family value system, and family problem-solving capability—with 5 items each, totaling 20 items. Each item was measured on a 4-point Likert scale (1: Not at all, 4: Always), where higher scores indicate a more positive state of family relationships.

- Dependent Variable: Social Relationship Satisfaction

Social relationship satisfaction was derived from the 'Social Relationships' item under the 'Current Life Satisfaction' section of the panel survey. This item captures the subjective level of satisfaction that respondents perceive regarding their current social networks. It is evaluated on a 10-point Likert scale, with higher scores representing superior satisfaction with social relationships.

### D. Data Analysis

Descriptive statistics and frequency analysis were performed to understand the general characteristics of the participants. Independent samples t-tests were conducted to compare family relationships and social relationship satisfaction between the child and adolescent groups. Additionally, multiple regression analysis controlling for general characteristics was executed to verify the effects of family relationships on social relationship satisfaction. Data preprocessing and statistical evaluations were carried out using R 4.5.1.

## III. Results

### A. General Characteristics of Participants

The participants consisted of 211 school-aged children and adolescents with developmental disabilities, including ID and ASD. The sample was composed of 148 children (70.10%) and 63 adolescents (29.90%), with a mean age of 10.99 years for the child group and 14.60 years for the adolescent group. Regarding sex, males numbered 126 (59.70%) and females numbered 85 (40.30%); the sex distribution did not differ significantly across school levels. In terms of disability type, ID accounted for 104 (49.30%) and ASD accounted for 107 (50.70%), showing a similar overall proportion.

### B. Comparison of Family Relationships and Social Relationship Satisfaction

The results of analyzing the differences in family relationships and social relationship satisfaction between the child and adolescent groups are presented in [Table 1]. The analysis demonstrated that total scores for the four sub-domains of family relationships—family cohesion, family communication, shared family value system, and family problem-solving capability—were slightly higher in the child group, but no statistically significant differences were confirmed at the domain level.

However, a detailed item-level analysis revealed significant discrepancies in certain areas. First, within the family cohesion domain, children (M=3.74, SD=0.46) scored significantly higher than adolescents (M=3.57, SD=0.56) in the 'Protection and Support' item. This suggests stronger perceived family support in childhood. Second, within the shared family value system domain, the child group (M=3.35, SD=0.78) also demonstrated a significantly higher score in 'Clarity of Rules and Values' than the adolescent group (M=3.10, SD=0.86). This suggests weaker rule clarity in adolescence as autonomy needs increase.

TABLE I
Comparison of Family Relationships by Developmental Stage

| Category | Children (n=148) | Adolescents (n=63) | *p* |
|---|---|---|---|

| | *M* | *SD* | *M* | *SD* | |
|---|---|---|---|---|---|
| **Family Cohesion** | 3.56 | 0.40 | 3.49 | 0.56 | >.35 |
| Family Harmony | 3.53 | 0.60 | 3.48 | 0.64 | 0.59 |
| Protection and Support | 3.74 | 0.46 | 3.57 | 0.56 | 0.04* |
| Understanding Emotions | 3.43 | 0.64 | 3.40 | 0.77 | 0.79 |
| Expressing Interest | 3.61 | 0.53 | 3.52 | 0.62 | 0.35 |
| Sincerity | 3.51 | 0.62 | 3.48 | 0.69 | 0.71 |
| **Family Communication** | 3.52 | 0.44 | 3.43 | 0.56 | 0.26 |
| Honesty among Members | 3.64 | 0.55 | 3.54 | 0.67 | 0.32 |
| Trust among Members | 3.50 | 0.69 | 3.41 | 0.73 | 0.42 |
| Participation in Decision-making | 3.45 | 0.63 | 3.33 | 0.72 | 0.28 |
| Listening | 3.46 | 0.64 | 3.40 | 0.77 | 0.15 |
| Freedom of Emotional Expression | 3.53 | 0.64 | 3.54 | 0.59 | 0.89 |
| **Shared Family Value System** | 3.42 | 0.47 | 3.26 | 0.66 | 0.09 |
| Consistency in Values and Beliefs | 3.41 | 0.67 | 3.32 | 0.74 | 0.42 |
| Consistency in Core Values | 3.50 | 0.59 | 3.46 | 0.64 | 0.15 |
| Possession of Family Traditions | 3.36 | 0.65 | 3.14 | 0.95 | 0.09 |
| Clarity of Rules and Values | 3.35 | 0.78 | 3.10 | 0.86 | 0.04* |
| Collaborative Rule-making | 3.47 | 0.62 | 3.41 | 0.75 | 0.58 |
| **Family Problem-Solving Capability** | 3.54 | 0.40 | 3.44 | 0.55 | 0.17 |
| Family-Centered Problem-Solving | 3.59 | 0.56 | 3.40 | 0.77 | 0.22 |
| Help-Seeking | 3.60 | 0.52 | 3.56 | 0.64 | 0.62 |
| Collaborative Problem-Solving | 3.59 | 0.55 | 3.46 | 0.64 | 0.17 |
| Diverse Problem-Solving Approaches | 3.55 | 0.63 | 3.53 | 0.63 | 0.89 |
| Resolving Differences | 3.39 | 0.68 | 3.21 | 0.83 | 0.13 |
| Social Relationship Satisfaction | **Children (*n*=148)** | | **Adolescents (*n*=63)** | | ***p*** |
| | ***M*** | ***SD*** | ***M*** | ***SD*** | |
| Social Relationship Satisfaction | 5.67 | 1.97 | 6.11 | 2.13 | 0.16 |

### C. *Effects of Family Relationships on Social Relationship Satisfaction*

In the multiple regression analysis for the entire sample, family communication ( $\beta$=2.14, $p$<.01) and family problem-solving capability ( $\beta$=-1.71, $p$<.01) exerted significant impacts on social relationship satisfaction. At the item level, honesty among family members ( $\beta$ = 0.90, $p$< 01) and participation in decision-making ($\beta$ = 0.65, $p$<.05) had positive effects, while help-seeking behaviors ( $\beta$ =-0.86, $p$<.01) showed a negative effect.

For the child group, only the domain of family communication ( $\beta$ =2.02, $p$< .01) significantly and positively predicted social relationship satisfaction, with honesty among family members ( $\beta$ =0.84, p< .05$) and participation in decision-making ( $\beta$ = 0.75, p <.05) standing out as core items. Cohesion, shared values, and problem-solving capabilities were non-significant.

For the adolescent group, family communication ( $\beta$ =2.27, $p$<.05) and trust among family members ( $\beta$ =1.97, p<.01) were identified as prominent factors that enhanced social relationship satisfaction. Conversely, family problem-solving capability ( $\beta$ =-2.57, p<.05) had a significant negative impact. This suggests that during adolescence, emotional trust and open communication are more important than direct parental intervention.

TABLE II
EFFECTS OF FAMILY RELATIONSHIPS ON SOCIAL RELATIONSHIP SATISFACTION

| Category | Children (*n*=148) | | | | Adolescents (*n*=63) | | | |
|---|---|---|---|---|---|---|---|---|
| | ***β*** | ***S.E*** | ***t*** | ***p*** | ***β*** | ***S.E*** | ***t*** | ***p*** |
| **Family Cohesion** | -0.25 | 0.66 | -0.37 | .71 | 1.64 | 0.86 | 1.9 | .06 |
| Family Harmony | 0.42 | 0.38 | 1.09 | .28 | -0.40 | 0.77 | -0.52 | .61 |
| Protection and Support | 0.36 | 0.42 | 0.86 | .39 | -0.05 | 0.93 | -0.06 | .96 |
| Understanding Emotions | -0.49 | 0.35 | -1.40 | .17 | 0.87 | 0.69 | 1.27 | .21 |
| Expressing Interest | -0.68 | 0.37 | -1.81 | .07 | 1.31 | 0.73 | 1.80 | .08 |
| Sincerity | 0.05 | 0.37 | 0.14 | .89 | -0.75 | 0.72 | -1.04 | .30 |
| **Family Communication** | 2.02 | 0.54 | 3.76 | .00 | 2.27 | 1.08 | 2.1 | .04* |
| Honesty among Members | 0.84 | 0.41 | 2.07 | .04 | 0.99 | 0.71 | 1.40 | .17 |
| Trust among Members | 0.35 | 0.32 | 1.10 | .28 | 1.97 | 0.64 | 3.08 | .00 |
| Participation in Decision-making | 0.75 | 0.32 | 2.36 | .02 | 0.27 | 0.77 | 0.36 | .72 |
| Listening | 0.14 | 0.34 | 0.42 | .67 | 0.73 | 0.83 | 0.89 | .38 |
| Freedom of Emotional Expression | 0.01 | 0.31 | 0.05 | .96 | -0.38 | 0.80 | -0.47 | .64 |
| **Shared Family Value System** | 0.17 | 0.48 | 0.36 | .72 | -0.02 | 0.69 | -0.03 | .98 |
| Consistency in Values and Beliefs | 0.27 | 0.33 | 0.81 | .42 | -0.64 | 0.65 | -0.98 | .33 |
| Consistency in Core Values | -0.23 | 0.35 | -0.65 | .51 | -0.16 | 0.60 | -0.26 | .80 |
| Possession of Family Traditions | -0.13 | 0.32 | -0.41 | .68 | -0.82 | 0.54 | -1.52 | .14 |
| Clarity of Rules and Values | 0.17 | 0.24 | 0.72 | .47 | 0.49 | 0.43 | 1.12 | .27 |
| Collaborative Rule-making | -0.21 | 0.36 | -0.58 | .56 | 0.89 | 0.82 | 1.10 | .28 |
| **Family Problem-Solving Capability** | -1.18 | 0.64 | -1.85 | .07 | -2.57 | 1.07 | -2.4 | .02* |
| Family-Centered Problem-Solving | 0.22 | 0.36 | 0.61 | .54 | -1.40 | 0.78 | -1.78 | .08 |
| Help-Seeking | -0.52 | 0.39 | -1.32 | .19 | -1.28 | 0.70 | -1.84 | .07 |
| Collaborative Problem-Solving | -0.51 | 0.39 | -1.32 | .19 | -0.30 | 0.78 | -0.38 | .70 |
| Diverse Problem-Solving Approaches | 0.18 | 0.34 | 0.54 | .59 | 0.21 | 0.65 | 0.32 | .75 |
| Resolving Differences | -0.16 | 0.30 | -0.55 | .58 | -0.27 | 0.49 | -0.54 | .59 |

## IV. CONCLUSIONS

This study utilized data from the 5th wave (2022) of the Disability and Life Dynamics Panel Survey to comparatively analyze the effects of family relationships on social relationship satisfaction between children and adolescents with DD [19]. The major conclusions drawn from the findings are as follows.

First, the baseline dynamics of family relationships differed by developmental stage; specifically, the child group perceived significantly higher levels of 'Protection and Support' and 'Clarity of Rules and Values' than the adolescent group. This confirms that the family during childhood acts as a primary emotional safety net and provides essential relational guideposts. The decreasing density of these elements during adolescence reflects a natural systemic shift driven by the child's expanding peer networks and intensifying autonomy

needs, which aligns with previous findings highlighting the necessity of restructuring family interactions during school-age transitions [17].

Second, the structural paths through which family relationships affect social relationship satisfaction were distinctively differentiated by developmental stage. In childhood, democratic participation and honest communication within the home foster interpersonal confidence, which successfully transfers into external social networks. This empirically supports the ecological perspective positioning the family as the primary microsystem [7] and resonates with the psychological mechanism of satisfying autonomy, competence, and relatedness needs asserted by self-determination theory [20]. In contrast, while 'Trust among members' serves as the psychological anchor during adolescence [13], parent-driven intervention styles (family problem-solving capability) backfired, exerting a negative effect. This reinforces previous assertions that proactive parental interventions can inadvertently restrict an adolescent's independent coping opportunities [15] and highlights that balancing family cohesion and adaptability is an essential equilibrium point for DD households under the Circumplex Model [21].

Third, these outcomes strongly emphasize the requirement for life-cycle-specific family support strategies. Family interventions for children with DD should prioritize enhancing open dialogue and scaling up opportunities for domestic decision-making [6]. For adolescents, however, there is an urgent need to transition parental roles from authoritative controllers to emotional facilitators, adopting a paradigm of 'Supportive Distancing' [15]. This does not imply detachment or neglect, but rather a democratic relationship style that respects the child's self-determination while providing supportive involvement. Operationally, regional support centers for individuals with DD must segment their counseling programs by life cycle [22], and school systems should establish reciprocal linkages with homes to ensure consistency between domestic parenting styles and school adjustment.

In conclusion, this study empirically established that the impacts of family relationships on social relationship satisfaction are substantively differentiated by developmental stage among students with DD. Moving forward, policy planners and clinical practitioners must systematically deploy customized family support frameworks tailored to specific life cycles, thereby driving successful community integration and maximizing life satisfaction for school-aged students with DD.

# Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176).

## REFERENCES

[1] *Act on Guarantee of Rights of and Support for Persons with Developmental Disabilities*, Law No. 18987, 2022.

[2] *Act on Guarantee of Rights of and Support for Persons with Developmental Disabilities*, Law No. 20182, 2024.

[3] Ministry of Education, *2025 Special Education Statistics*, Sejong: Ministry of Education, 2025.

[4] C. J. Dunst, "Family-centered practices: Birth through high school," *The Journal of Special Education*, vol. 36, no. 3, pp. 141-149, 2002.

[5] P. Rosenbaum, "Childhood developmental disability: Recognising the primary role of the family," *BMJ Paediatrics Open*, vol. 9, no. 1, e003730, 2025.

[6] M. J. Han, "A study on family life experiences and family support needs of parents of children with developmental disabilities," *Journal of Intellectual Disabilities*, vol. 27, no. 4, pp. 115-144, 2025.

[7] U. Bronfenbrenner, *The Ecology of Human Development: Experiments by Nature and Design*, Harvard University Press, 1979.

[8] Z. Amanollahi, F. GhasemiNiaei, F. F. Davarani, and M. M. Esfahan, "Development of an integrated developmental model for adaptation with autistic children in family system," Journal of Autism and Developmental Disorders, pp. 1-14, 2025.

[9] C. I. Iacob, E. Avram, D. Cojocaru, and I. R. Podina, "Resilience in familial caregivers of children with developmental disabilities: A meta-analysis," *Journal of Autism and Developmental Disorders*, vol. 50, no. 11, pp. 4053-4068, 2020.

[10] Y. Yoo, "Development of the Korean family strength scale for family empowerment," *Korean Journal of Family Relations*, pp. 119-151, 2004.

[11] E. J. Eo and Y. Yoo, "A study on the development of a scale for measuring 'family strength'," *Journal of Korean Home Management Association*, vol. 13, no. 1, pp. 145-156, 1995.

[12] J. H. Choi, "A study on family stress, coping strategies and family strength of weekend couples," *Journal of Korean Home Management Association*, vol. 22, no. 2, pp. 69-83, 2004.

[13] H. Shin and M. H. Jung, "The effect of family resilience on social adaptation of children with developmental disabilities: Mediating effects of parenting efficacy and coping strategies," *Journal of the Next-generation Convergence Technology Association*, vol. 9, no. 10, pp. 2610-2618, 2025.

[14] T. Y. Shim, "The effect of a family-centered camping program on social skills and subjective happiness of students with developmental disabilities," *Korean Journal of Sport Studies*, vol. 23, no. 3, pp. 239-249, 2025.

[15] S. H. Lee, J. H. Hong, J. H. Yeom, and J. A. Lee, "A preliminary study on the development of self-determination instructional support programs at home: Focusing on experiences and support needs of parents and teachers of students with disabilities," *Korean Journal of Special Education*, vol. 54, no. 1, pp. 25-60, 2019.

[16] B. G. Byun and E. S. Jeon, "A study on parenting stress and family support of parents of children with disabilities," *Theory and Practice of Education*, vol. 11, no. 2, pp. 71-88, 2006.

[17] S. R. Kwak and K. H. Lee, "Causal relationships among variables related to school adjustment of students with disabilities in inclusive elementary classrooms," *Journal of Developmental Disabilities*, vol. 26, no. 4, pp. 485-507, 2022.

[18] J. E. Choi, "The effect of family strength on life satisfaction through disability acceptance based on individual analysis of persons with developmental disabilities," *Journal of Developmental Disabilities*, vol. 29, no. 2, pp. 225-240, 2025.

[19] Korea Disabled People's Development Institute, *Disability and Life Dynamics Panel Survey User Guide*, Seoul: Korea Disabled People's Development Institute, 2022.

[20] E. L. Deci and R. M. Ryan, "The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior," *Psychological Inquiry*, vol. 11, no. 4, pp. 227–268, 2000.

[21] D. H. Olson, "Circumplex model of marital and family systems," *Journal of Family Therapy*, vol. 22, no. 2, pp. 144–167, 2000.

[22] H. S. Park, "Family support plans for families of children with autism spectrum disorders," *Special Education*, vol. 5, no. 2, pp. 5-30, 2006.

[23] K. H. Lee, "The effect of family strength and disability impact on school adjustment of elementary students with disabilities in inclusive classrooms," *Journal of Inclusive Education*, vol. 17, no. 2, pp. 1-20, 2022.

# A Comprehensive Preprocessing Framework for Automated Handwriting Evaluation Aligned with the Minnesota Handwriting Assessment

**J.T. Amores,[1,2] A.C. Frias,[1,2] P.M. Espino,[3] I.B. Gomez,[3] and E.A. Roxas[2]**

*[1]Graduate School, University of Santo Tomas, España, Manila, Philippines*
*[2]Faculty of Engineering, University of Santo Tomas, España, Manila, Philippines*
*[3]College of Rehabilitation Sciences, University of Santo Tomas, España, Manila, Philippines*
**Corresponding author: josephden.amores.gs@ust.edu.ph*

Accurate handwriting assessment in children is essential for evaluating fine motor skills, cognitive development, and academic readiness. Standardized tools such as the Minnesota Handwriting Assessment (MHA) provide structured criteria for evaluating legibility, form, alignment, size, and spacing, yet manual scoring is subjective, time-consuming, and challenging to scale across large and marginalized populations. This study presents a comprehensive preprocessing framework that serves as a foundation and offers support for automated handwriting evaluation directly aligned with MHA standards. Handwritten samples from children aged 6–10 years were captured using a reproducible prototype comprising an LED-illuminated studio softbox and camera, ensuring consistent orientation, lighting, and high-quality image acquisition across diverse participants. The preprocessing pipeline includes region-of-interest cropping, grayscale conversion, non-local means denoising, adaptive thresholding, guideline removal, contour-based noise filtering, and morphological refinement, producing clean, high-contrast images of individual letters. Interactive letter cropping ensures precise isolation of each character, preserving structural integrity critical for evaluating MHA-specific handwriting components. Preliminary results demonstrate that the pipeline effectively reduces background noise, compensates for illumination and ink variations, and maintains fine motor details necessary for scoring MHA criteria. By producing standardized letter-level inputs, this framework provides a foundation for scalable and objective automated evaluation, enabling future systems to deliver interpretable, component-wise assessments consistent with MHA scoring guidelines. This methodology provides a novel, fundamental, practical approach for enhancing accessibility, objectivity, and efficiency in educational and clinical handwriting assessment.

# EXPLORING THE INTEGRATION OF ARTIFICIAL INTELLIGENCE TOOLS TO ENHANCE COUNSELING PRACTICES AMONG FILIPINO OCCUPATIONAL THERAPISTS

J.G. Cua, M.S. Tan[1], G.B. Evangelista[2], M.L. Chavez[2]

*Department of Occupational Therapy, University of Santo Tomas, Manila, Philippines* [1]

*Department of Psychology, Miriam College, Manila, Philippines* [2]

*Department of Education, Miriam College, Manila, Philippines* [3]

**Corresponding author: jcgurtiza@ust.edu.ph*

Artificial intelligence (AI) is increasingly utilized in healthcare; however, its integration into occupational therapy (OT) counseling remains insufficiently examined within the Philippine context. This study investigated the ethical and contextual integration of AI into OT counseling practice and education through a Participatory Action Research approach. Data collection involved a focus group discussion with licensed occupational therapists and five key informant interviews with educators, clinicians, and a school-based guidance counselor. Participant-validated transcripts were analyzed using Reflexive Thematic Analysis, following Braun and Clarke's six-phase framework. The analysis identified five themes: AI awareness, barriers and enablers to integration, applications in counseling-oriented OT practice, and recommendations for ethical use. Results suggest that Filipino occupational therapists possess emerging yet limited familiarity with AI, coupled with a strong interest in structured guidance and safeguards. The study developed an AI Protocol Manual to facilitate ethical and culturally responsive integration of AI in OT counseling, underscoring that AI should augment rather than replace the human-centered foundations of practice.

**Keywords:** artificial intelligence, occupational therapy, counseling, ethics, education, participatory action research, Philippines

## Methodology

***Figure 1***

Participatory Action Research Framework for the Development of an AI Protocol for Filipino Occupational Therapy




## Results

***Figure 2***

**Conceptual Map**

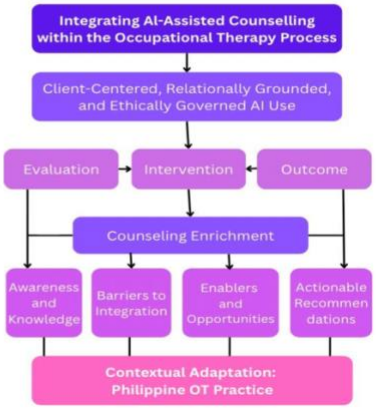



*Figure 2* serves as a visual synthesis of the study's findings, illustrating that the Integration of AI into occupational therapy must occur within a cyclical and relational process anchored in counseling. Rather than a linear or mechanistic tool, AI operates within and across the stages of the OT process. During evaluation, AI supports information gathering, reflection, and data synthesis. Within the intervention, it assists in generating counseling resources, tracking client progress, and structuring reflective dialogue. In outcomes, it enables monitoring, documentation, and adaptive planning. Surrounding this core cycle are the five interrelated themes, each representing a developmental condition that enables ethical and contextual Integration. Together, these components form a reflexive ecosystem of augmented practice—a model where echnology, therapist judgment, and client participation co-evolve to strengthen the therapeutic alliance and promote culturally grounded, client-centered care.

## Implications

### Implications for Counseling Practice in Occupational Therapy

This study reaffirms that counseling is not a supplementary activity in occupational therapy (OT) but a central, relational process that sustains empathy, ethical reflection, and client-centered care, and integrating artificial intelligence (AI) into counseling-oriented OT practice requires a reframing of technology's purpose—from a mechanistic substitute for reasoning to a reflective partner that enhances therapeutic presence and attunement. Participants consistently described AI as a cognitive aid that helps organize thoughts, summarize reports, and manage documentation, allowing therapists to devote more time and attention to the relational aspects of client care. However, the value of AI integration depends on maintaining counseling ethics as its foundation. Practitioners emphasized that empathy, presence, and relational integrity cannot be automated; they require genuine human interaction. Counseling principles—trust, authenticity, and mutual respect—must continue to guide the therapist's reflective reasoning. In this sense, AI functions as a cognitive amplifier, enhancing the therapist's ability to reflect, plan, and evaluate, but never replacing the human process of therapeutic understanding. Also, the use of AI must therefore remain transparent, ethically supervised, and guided by professional discernment to ensure that the therapist's presence remains at the core of the therapeutic alliance.

### Implications for Model Building

The study contributes to model building by proposing an integrative conceptual framework that positions counseling at the center of the OT process while embedding AI as a supportive and adaptive layer. The model illustrated in Figure 2: *Conceptual Integration Model*, demonstrates how counseling bridges the phases of the Occupational Therapy Practice Framework (AOTA, 2020)—evaluation, intervention, and outcomes—while AI serves as an assistive layer that supports reflection, ethical reasoning, and clinical decision-making. The model reframes counseling as a cross-cutting mechanism in the OT process rather than a distinct intervention category. It functions as the interpretive and relational thread that connects data with human meaning. The findings suggest that AI can operationalize what Topol (2019) refers to as *augmented intelligence*—a form of collaboration between human insight and machine efficiency that enhances, rather than replaces, professional judgment. Within this paradigm, AI's role is reflective and supportive, not directive or deterministic. The Filipino adaptation of the model contextualizes global AI ethics within local values of *pakikipagkapwa* (shared humanity) and *bayanihan* (collective responsibility). These cultural foundations affirm that technological adoption must preserve the relational and communal orientation of care that defines Filipino counseling and OT practice. Model development,

therefore, should continue to integrate cultural sensitivity and ethical literacy as operational dimensions of AI-assisted occupational therapy. This integrative framework can guide the development of future protocols, curriculum reforms, and practice guidelines that reflect both global standards and local contexts, thereby enhancing the effectiveness of healthcare delivery.

**Implications for Occupational Therapy Education**

There is a pressing need to modernize occupational therapy education in response to the digital transformation in health and rehabilitation. The current CHED Memorandum Order (CMO) No. 52, Series of 2017, defines program outcomes related to ethical reasoning and reflective practice but omits competencies related to AI, digital health, and data literacy. Participants' reflections show that most practitioners and educators learned about AI independently, outside formal instruction, often through self-exploration and informal experimentation. Integrating AI literacy into OT curricula must go beyond technical instruction. It should emphasize the reflective and ethical use of technology in counseling contexts. Embedding digital ethics, simulation-based learning, and case-based reflection into psychosocial and counseling-related courses can enhance students' readiness to navigate complex, technology-assisted therapeutic scenarios. Interdisciplinary collaboration among OT, psychology, computer science, and bioethics faculties can further strengthen the educational infrastructure for the responsible use of AI. Faculty development and continuing professional education are equally vital. Continuing Professional Development (CPD) programs should include training in AI ethics, data governance, and applied counseling supervision to ensure practitioners remain adaptable and ethically grounded throughout their careers. This approach aligns with Knowles' (1984) principles of adult learning, which emphasize self-directed, experiential, and contextually relevant learning as essential to professional growth.

**Implications for Policy and Institutional Governance**

At the policy level, the findings underscore the need for institutional frameworks that ensure the integration of AI remains ethical, transparent, and contextually appropriate. Professional bodies such as the Philippine Academy of Occupational Therapy (PAOT) and the Professional Regulation Commission (PRC) are in strategic positions to establish national guidelines that regulate the ethical use of AI in OT practice. Participants repeatedly called for formal policies, emphasizing that "We need clear guidelines on how to use AI ethically" (KII Participant, 2025).

Policy initiatives should address data privacy, algorithmic accountability, and informed consent, in line with the Philippine Data Privacy Act of 2012 (RA 10173) and the World Health Organization's (2021) principles for AI governance. Health and academic institutions can establish AI Ethics Review Boards to monitor implementation, assess risk, and guide ethical decision-making in practice and education. Additionally, AI-related competencies can be embedded in accreditation, licensure, and CPD frameworks, ensuring that ethical and technological literacy become hallmarks of professional competence. Institutional partnerships with technology developers and universities should be encouraged to co-create culturally responsive AI tools that align with Filipino psychosocial contexts. This collaborative, multi-sectoral approach can help strike a balance between innovation and ethical stewardship, ensuring that technological advancements serve the greater good of human well-being.

**Implications for Further Research**

The study identifies several directions for future research on integrating artificial intelligence (AI) into counseling-oriented occupational therapy (OT). Building on current qualitative findings and the development of the AI protocol manual, subsequent research should progress to Phase 6: Implementation and Evaluation, which remains unaddressed in this investigation. Future studies should examine the real-world application of the protocol manual in both academic and clinical OT settings through rigorously designed pilot implementations. These studies may employ quantitative and mixed-methods approaches to assess the feasibility, adaptability, and outcomes of AI-supported counseling practices.

When evaluating implementation, future research should consider metrics such as practitioner adherence to the protocol's ethical guidelines, including consistent use of AI-specific informed consent procedures and completion of ethical checklists before AI-supported sessions. Process indicators may include the frequency and type of AI-supported activities adopted, such as documentation support, psychoeducational material generation, and counseling preparation. Additional measures may include changes in time allocation between administrative tasks and direct client interaction, as well as practitioner self-reported confidence and ethical comfort in using AI-assisted tools. These indicators can clarify whether AI integration meaningfully supports counseling practice without displacing professional judgment or relational engagement.

At the client level, future research should incorporate evaluation measures related to counseling quality and relational outcomes. These measures may include client-reported understanding of AI use, perceived transparency and trust, comfort with AI-supported materials, and perceived impact on engagement and participation in therapy. Importantly, consistent with the study's Participatory Action Research (PAR) orientation, clients may be actively involved in co-defining what transparency and trust mean in AI-assisted counseling contexts, rather than treating these constructs solely as predefined variables. Through collaborative feedback sessions, reflective interviews, or participatory evaluation activities, clients can articulate the indicators of ethical clarity, relational safety, and openness that are most meaningful to them. Where appropriate, brief assessments of therapeutic alliance stability or qualitative feedback on relational safety may be used to examine whether AI integration maintains the counseling values central to occupational therapy practice. Collectively, these collaboratively informed metrics would enable empirical testing of the protocol manual's underlying assumptions and demonstrate its practical and relational value across diverse OT environments.

Further research should address Phase 7: Dissemination and Policy Recommendations, which was intentionally beyond the scope of the current study. This phase may involve distributing the finalized AI protocol manual to academic institutions, professional organizations, and policymakers, followed by systematic examination of its influence on curriculum development, faculty training, continuing professional development programs, and ethical governance related to AI use in OT counseling. Policy-oriented studies should assess institutional uptake of protocol recommendations, alignment with national educational and professional standards, and the establishment of governance mechanisms for AI vetting, data privacy, and ethical oversight.

Incorporating client and family perspectives is essential for understanding how AI-mediated counseling practices influence perceptions of empathy, trust, and relational safety. Additional research focusing on educator readiness, institutional capacity, and resource availability can further inform curriculum reforms that integrate AI ethics, reflective counseling practice, and culturally responsive competencies. Consistent with the study's Participatory Action Research (PAR) foundation, future investigations should co-design, implement, and evaluate AI-supported counseling practices with Filipino occupational therapists, educators, and clients to ensure cultural, linguistic, and ethical relevance.

References

Abd-Alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, *132*, 103978. https://doi.org/10.1016/j.ijmedinf.2019.103978

Aguas, J. J. S. (2016). The Filipino Value of Pakikipagkapwa-Tao Vis-À-Vis Gabriel Marcel's Notion of Creative Fidelity and Disponibilitè. *Scientia: Research Journal of the College of Arts and Sciences, San Beda University*.

Ahuja, A. S. (2019). The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ*, *7*, e7702. https://doi.org/10.7717/peerj.7702

Aljerian, N., Arafat, M., Aldhubib, A., Almohaimeed, I., Alsultan, A., Alhosaini, A., Aloqayli, A., AlRabiah, B., AlBanyan, S., Ashry, L. B., AlGhassab, R., Alzahrani, R., Alhenaki, G., Alnazzawi, E., AlMughaiseeb, A., & Alanazi, A. (2022). Artificial

Intelligence in Health care and its application in Saudi Arabia. *International Journal of Innovative Research in Medical Science*, *7*(11), 666–670. https://doi.org/10.23958/ijirms/vol07-i11/1558

American Occupational Therapy Association. (2020). Occupational therapy practice framework: Domain and process (4th ed.). *American Journal of Occupational Therapy*, *74*(Suppl. 2), 7412410010. https://doi.org/10.5014/ajot.2020.74S2001

Aponte, H. J. (2021). The soul of Therapy: the therapist's use of self in the therapeutic relationship. *Contemporary Family Therapy*, *44*(2), 136–143. https://doi.org/10.1007/s10591-021-09614-5

Arushi, Dillon, R., Teoh, A. N., & Dillon, D. (2022). Voice Analysis for stress detection and application in virtual Reality to improve public speaking in real-time: A review. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2208.01041

Avello-Sáez, D., & Estrada-Palavecino, L. (2023). ChatGPT and its impact on competence training in occupational therapists: a reflection on academic integrity. *Cadernos Brasileiros de Terapia Ocupacional*, *31*. https://doi.org/10.1590/2526-8910.ctoen271035342

B, N. P. K., B, N. S., & Patrer, N. D. (2020). Emotion recognition with AI: Techniques and applications. *World Journal of Advanced Research and Reviews*, *8*(2), 344–352. https://doi.org/10.30574/wjarr.2020.8.2.0362

Bankins, S., Hu, X., & Yuan, Y. (2024). Artificial intelligence, workers, and future of work skills. *Current Opinion in Psychology*, *58*, 101828. https://doi.org/10.1016/j.copsyc.2024.101828

Basu, K., Sinha, R., Ong, A., & Basu, T. (2020). Artificial intelligence: How is it changing medical sciences and its future? *Indian Journal of Dermatology*, *65*(5), 365. https://doi.org/10.4103/ijd.ijd_421_20

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage High-Risk and High-Cost patients. *Health Affairs*, *33*(7), 1123–1131. https://doi.org/10.1377/hlthaff.2014.0041

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

https://doi.org/10.1191/1478088706qp063oa Chandler, D., & Torbert, B. (2003). Transforming inquiry and action. *Action Research*, *1*(2), 133–152. https://doi.org/10.1177/14767503030012002

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, *4*(2), e19. https://doi.org/10.2196/mental.7785

Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). SAGE Publications.

Garcia, J. A., Lorenzo, F. M., & David, E. J. R. (2021). Occupational therapy in the Philippines: Challenges and opportunities in post-disaster rehabilitation. *Philippine Journal of Occupational Therapy*, *15*(1), 12-25.

Gutierrez, K. L. T., & Viacrusis, P. M. L. (2023). Bridging the gap or widening the divide: A call for Capacity-Building in artificial intelligence for healthcare in the Philippines. *Journal of Medicine University of Santo Tomas*, *7*(2), 1325–1334. https://doi.org/10.35460/2546-1621.2023-0081

Habermann, J. (2021). *Psychological Impacts of COVID-19 and Preventive Strategies: A review*. Humanities Publishing House. https://doi.org/10.13140/RG.2.2.29662.15681

Haglund, L., Thorell, L. H., & Walinder, J. (2000). Occupational therapy and the therapeutic use of self. *Scandinavian Journal of Occupational Therapy*, *7*(1), 11-16. https://doi.org/10.1080/110381200443572

Haitham Jahrami. (2023). Unlocking the potential: Exploring the opportunities for occupational therapists in cyberpsychology. *Hong Kong Journal of Occupational Therapy*. https://doi.org/10.1177/15691861231222074

Hisler, G. C., Young, K. S., Cumpanasoiu, D. C., Palacios, J. E., Duffy, D., Enrique, A., Keegan, D., & Richards, D. (2024). Incorporating a deep-learning client outcome

prediction tool as feedback in supported internet-delivered cognitive

behavioural therapy for depression and anxiety: A randomised controlled trial within routine clinical practice. *Counselling and Psychotherapy Research*. https://doi.org/10.1002/capr.12771

Kinsella, E. A. (2020). Professional identity in occupational therapy: A constructivist grounded theory study. Scandinavian Journal of Occupational Therapy, 27(4),

251-262. https://doi.org/10.1080/11038128.2019.1577168
Iannone, A., & Giansanti, D. (2023). Breaking Barriers—The Intersection of AI and

Assistive Technology in Autism Care: A Narrative review. *Journal of Personalized Medicine*, *14*(1), 41. https://doi.org/10.3390/jpm14010041 Johnson, D., Deterding, S., Kuhn, K. A., Staneva, A., Stoyanov, S., & Hides, L. (2016).

Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions*, *6*, 89-106. https://doi.org/10.1016/j.invent.2016.10.002

Kinsella, E. A. (2020). Professional identity in occupational therapy: A constructivist grounded theory study. *Scandinavian Journal of Occupational Therapy*, *27*(4), 251-262. https://doi.org/10.1080/11038128.2019.1577168

Lima, R., Chirico, A., Varandas, R., Gamboa, H., Gaggioli, A., & Badia, S. B. I. (2024). Multimodal emotion classification using machine learning in immersive and non-immersive virtual reality. *Virtual Reality*, *28*(2). https://doi.org/10.1007/s10055-024-00989-y

Lorenzo, F. M., Garcia, J. A., & Enriquez, V. G. (2018). Bayanihan in occupational therapy: Community-based rehabilitation in the Philippines. *Asian Journal of Occupational Therapy*, *14*(2), 45-52.

Luxton, D. D. (2020). Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization*, *98*(4), 285-287. https://doi.org/10.2471/BLT.19.237636

Miner, A. S., Milstein, A., & Hancock, J. T. (2017). Talking to machines about personal mental health problems. *JAMA*, *318*(13), 1217. https://doi.org/10.1001/jama.2017.14151

Naderbagi, A., Loblay, V., Zahed, I. U. M., Ekambareshwar, M., Poulsen, A., Song, Y. J. C., Ospina-Pinillos, L., Krausz, M., Kamel, M. M., Hickie, I. B., & LaMonica, H. M. (2024). Cultural and Contextual adaptation of Digital Health Interventions: Narrative review. *Journal of Medical Internet Research*, *26*, e55130. https://doi.org/10.2196/55130

Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., & Eberhardt, J. (2024). Enhancing mental health with Artificial Intelligence: Current trends and future prospects. *Journal of Medicine Surgery and Public Health*, *3*, 100099. https://doi.org/10.1016/j.glmedi.2024.100099

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., ... & UnÜtzer, J. (2020). The Lancet Commission on global mental health and sustainable development. The Lancet, 392(10157), 1553-1598. https://doi.org/10.1016/S0140-6736(18)31612-X

Parsons, T. D., & Rizzo, A. A. (2007). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(3), 250–261. https://doi.org/10.1016/j.jbtep.2007.07.007

Petersson, L., Larsson, I., Nygren, J., Nilsen, P., Neher, M., Reed, J., Tyskbo, D., & Svedberg, P. (2022). Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. *BMC Health Services Research*, *22*(1). https://doi.org/10.1186/s12913-022- 08215-8

Reason, P., & Bradbury, H. (2008). *The Sage handbook of action research : participative inquiry and practice*. In Sage eBooks. https://jprm.scholasticahq.com/article/13244-participatory-research-methods- choice-point s-in-the-research-process

Reason, P., & Torbert, W. (2001). The action turn. *Concepts and Transformation*, *6*(1), 1– 37. https://doi.org/10.1075/cat.6.1.02rea

Reddy, Sandeep. (2021). Artificial intelligence and healthcare—why they need each other?. *Journal of Hospital Management and Health Policy*, *5*. 10.21037/jhmhp-2020-ai-03.

Reyes, J. (2015). LoóbandKapwa: An Introduction to a Filipino Virtue Ethics. *Asian Philosophy*, *25*(2), 148–171.

https://doi.org/10.1080/09552367.2015.1043173

Rogers, C. R. (1951). *Client-centered Therapy*. Boston: Houghton Mifflin.
Ross, M. K., Wei, W., & Ohno-Machado, L. (2014). “Big data” and the electronic health

record. *Yearbook of Medical Informatics*, *23*(01), 97–104.

https://doi.org/10.15265/iy-2014-0003

Schreier, M. (2012). Qualitative content analysis in practice. SAGE Publications. Soriano, C. R., David, C. C., & Torres, M. T. (2021). Trust in technology: A study of

Filipino attitudes toward AI in healthcare. *Philippine Journal of*

*Psychology*, *54*(1), 78-95.
Tanaka, G., Inadomi, H., Kikuchi, Y., & Ohta, Y. (2020). Evaluating an artificial

intelligence-powered mental health app in Japan: A feasibility study. Asian Journal of Psychiatry, 52, 102025. https://doi.org/10.1016/j.ajp.2020.102025

Taylor, R. R., Lee, S. W., Kielhofner, G., & Ketkar, M. (2009). Therapeutic use of self: A nationwide survey of practitioners’ attitudes and experiences. *American Journal of Occupational Therapy*, *63*(2), 198-207. https://doi.org/10.5014/ajot.63.2.198

114

Topol, E. J. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.

Vaughn, L. M., & Jacquez, F. (2020). Participatory Research Methods – Choice Points in the Research Process. *Journal of Participatory Research Methods*, *1*(1). https://doi.org/10.35844/001c.13244

Witty, M. C. (2006). Client-Centered therapy. In *Springer eBooks* (pp. 35–50). https://doi.org/10.1007/978-0-387-29681-4_3

World Health Organization. (2020). *Global strategy on human resources for health: Workforce 2030*. WHO Press.

**PRESENTATION TYPE**

☐ Oral Presentation
✔ Poster Presentation

**CONFERENCE TOPIC AREA**

☐ Social Inequality & Economic Disparities
☐ Public Health & Well-being
✔ Education & Social Change
☐ Environmental & Sustainability Issues
☐ Technology & Society
☐ Conflict Resolution & Human Rights
☐ Governance & Policy Making
☐ Artificial Intelligence
☐ Wireless Technologies
☐ Cyber Security
☐ Sustainable Technology & Emerging Technologies
☐ Other (please specify): _____

# Practices Of Filipino Speech-Language Pathologists Based In Metro Manila In Managing Pragmatic Difficulties Among Children With Autism Spectrum Disorder

**Bautista, Elinor C.; Esguerra, Mary Ann Gisselle O.; Mabanag, Maria Kiana; Adriatico, Anekiel Jayne; Dulay, Bianca Laine; Medallo, Angela Kyle; Tumang, Alexandra Nicole**

*Department of Speech-Language Pathology, University of Santo Tomas, Manila, Philippines*
*College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*
** Corresponding author: elcbautista@ust.edu.ph*

Pragmatic language skills are essential for meaningful social communication yet remain particularly challenging for children with Autism Spectrum Disorder (ASD).Despite the growing prevalence of ASD in the Philippines, no published research exists on the standards of practice for pragmatic assessment and intervention among Filipino speech-language pathologists (SLPs). This study aimed to describe the practices of Filipino SLPs in private clinics within the National Capital Region (NCR) in managing pragmatic language difficulties of school-aged children with ASD. A qualitative descriptive design was employed. Eight registered Filipino SLPs with at least five years of clinical experience participated in one-on-one semi-structured interviews conducted online. Data were analysed through Braun and Clarke's Six-Phase Thematic Analysis using NVivo software. Four major themes emerged: assessment, intervention planning, implementation, and collaborative efforts. Participants reported utilising formal assessment tools primarily to evaluate prerequisite receptive, expressive, and higher language skills, while informal and criterion-referenced approaches were preferred for assessing pragmatics due to the absence of culturally appropriate norms for Filipino children. Intervention planning was highly individualized and guided by the child's language abilities, interests, and social environment. Commonly utilised approaches included Social Thinking and peer-mediated intervention, although the need for cultural adaptation of Western-based frameworks was noted. Therapy sessions typically began through individualized sessions before transitioning to dyad or group settings using real-life, context-driven activities. Intervention activities included roleplay, social stories, videos, board games, and context-based conversational tasks designed to reflect real-life situations. Interdisciplinary collaboration with teachers, occupational therapists, and parents was considered essential for goal-setting, behavioural support, and carryover across settings. The findings suggest that Filipino SLPs' practices in managing pragmatic difficulties are context-driven, culturally responsive, and grounded in clinical reasoning, underscoring the need for locally developed assessment frameworks and intervention guidelines for children with ASD in the Philippine context.

**Keywords:** Practices, Autism Spectrum Disorder (ASD), Filipino Speech-Language Pathologists (SLPs), managing pragmatic difficulties, private clinics, Metro Manila

# Children's Screen Time and Its Correlation with Well-being, Parental Attitudes, and Guardians' Screen Time in Filipino Children in Malabon City: An Observational Cross-Sectional Study

**Espino, Paulin Grace M.[1,2]; Tuazon, John Alexander David B.[2]; Mortos, Mikaela Kassy O.[2]; Anteojo, Nashley Gayle DC.[2]; Nobleza, Kervin Joshua O.[2]; Pangan, Cestel Margaret R.[2]; Precioso, Maria Pauline G.[2]; Vargas, Marianne C.[2]**

[1] Center for Health Research and Movement Sciences, University of Santo Tomas, Manila, Philippines
[2] Department of Occupational Therapy, College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines
***** *Corresponding author: pmespino@ust.edu.ph*

The global trend toward increased screen use has led to a surge in screen time among young children, prompting attention to its relationship with children's well-being. Well-being includes physical, emotional, psychological, and social health. In the Philippines, where access to technology is growing rapidly, it is important to examine how it affects young children. This study aimed to: (1) determine the correlation between screen time and the well-being of young children, (2) identify the correlation between parental attitudes and children's screen time, and (3) identify the correlation between children's and their guardians' screen time. A cross-sectional, observational, and analytical research design was used in Filipino children in Kindergarten to Grade 2 enrolled in schools in Malabon City, Metro Manila. The Media Exposure Survey and the Filipino Child Well-Being were used to measure the study's variables. The Mann–Whitney U test was used to compare public and private school variables. A Spearman's rank-order correlation was performed to assess the relationship between the variables. No significant correlation was observed between children's screen time and well-being, or parental attitudes and children's screen time. A strong positive correlation was found between children's and guardians' time spent on screens ($r_s = 0.67$, $p < .001$). A very strong correlation between children's and guardians' screen time was found for Kindergarten students ($r_s = 0.80$, $p < .001$). The results indicate that the amount of time spent on screen use is not sufficient to determine correlation with child well-being and parental attitudes, implying the importance of screen use moderation rather than avoidance. The strong correlation between children's and their guardians' screen time suggests the importance of modeling in children's screen use. Understanding factors that contribute to guardians' excessive screen use remains important for future research, particularly in identifying supports that can help families develop better screen habits.

**Keywords:** Screen time, Child well-being, Parental attitudes, Screen type, Filipino children

# iSULAT: Feature Definition and Validation of a Sensor-Based Pen for Handwriting Analysis

A.C. Frias,[1,2] J.T. Amores,[1,2] P.M. Espino,[3] I.B. Gomez,[3] and E.A. Roxas[2]

[1]*Graduate School, University of Santo Tomas, España, Manila, Philippines*

[2]*Faculty of Engineering, University of Santo Tomas, España, Manila, Philippines*

[3]*College of Rehabilitation Sciences, University of Santo Tomas, España, Manila, Philippines*

**Corresponding author: antoniomiguel.frias.gs@ust.edu.ph*

Handwriting is a fundamental fine motor skill in early childhood education, with school-aged children spending a substantial portion of their academic activities engaged in writing-related tasks. Despite its importance, handwriting difficulties are reported in a considerable proportion of school-aged children, often affecting academic performance, confidence, and learning development. Existing handwriting assessment approaches rely heavily on expert observation and structured evaluation procedures, which can be time-intensive and limit opportunities for continuous and objective measurement. These challenges highlight the need for supportive tools that enhance and complement clinical assessment practices. This study presents the feature definition and validation of a sensor-based pen system, iSULAT (Intelligent Stroke Utilization, Learning, Assessment, and Testing), where “isulat” is the Filipino term for “to write”, reflecting the system’s focus on handwriting-centered analysis. The system integrates motion, pressure, and orientation sensors to acquire raw handwriting signals, from which thirteen handwriting features are formally defined, encompassing spatial, temporal, kinematic, pressure-related, and orientation-based characteristics of handwriting execution. Feature validity is established through multiple complementary approaches, including construct-based analysis, reliability assessment across repeated trials, expert validation by clinicians in the field of rehabilitation sciences, and verification using an independent camera-based recording setup to confirm temporal and spatial measurement accuracy. Results demonstrate stable feature behavior, expected inter-feature relationships, and consistency between sensor-derived measurements and observable writing events. By establishing a validated and interpretable set of handwriting features, this work provides a strong foundation for objective handwriting analysis and supports the development of technology-assisted tools that augment handwriting assessment and monitoring.

**PRESENTATION TYPE**

- [x] Oral Presentation
- [ ] Poster Presentation

**CONFERENCE TOPIC AREA**

- [ ] Social Inequality & Economic Disparities
- [ ] Public Health & Well-being
- [ ] Education & Social Change
- [ ] Environmental & Sustainability Issues
- [ ] Technology & Society
- [ ] Conflict Resolution & Human Rights
- [ ] Governance & Policy Making
- [ ] Artificial Intelligence
- [ ] Wireless Technologies
- [ ] Cyber Security
- [x] Sustainable Technology & Emerging Technologies
- [ ] Other (please specify): _____

# Development of a Thermo-Tactile Stimulation Device with EEG-based Neurofeedback for Stress Intervention

**Ocenar, Samuel John A.**[1,2,*]; **Fabay, Lianna Mae A.**[1,2]; **Danganan, Lester Mico L.**[5]; **Nieles, Jay Patrick M.**[1]; **Catindig-Dela Rosa, Joseree Ann S.**[6,7]; **Prado, Seigfred V.**[1,2,3,4]

[1] Brain Research Advancement and Innovations in Neurotechnology (BRAIN) Lab, University of Santo Tomas, Manila, Philippines

[2] Graduate School, University of Santo Tomas, Manila, Philippines

[3] Department of Electronics Engineering, University of Santo Tomas, Manila, Philippines

[4] Research Center for Engineering and Technological Sciences, University of Santo Tomas, Manila, Philippines

[5] NSW Health, Sydney, New South Wales, Australia

[6] Faculty of Medicine and Surgery, University of Santo Tomas, Manila, Philippines

[7] National Center for Geriatric Health, Manila, Philippines

* *Corresponding author: samueljohn.ocenar.gs@ust.edu.ph*

Stress is a pervasive concern among young adults aged 18 to 25, with university students particularly vulnerable due to academic, social, and financial pressures. Chronic stress impairs emotional regulation and cognitive function, increasing the risk for anxiety and depression. Although pharmacological and cognitive-behavioral interventions are available, their use may be limited by side effects, accessibility barriers, and high cognitive demands. Sensory-based interventions offer a low cognitive demand alternative by engaging peripheral pathways involved in relaxation. Thermal stimulation provides a gradual soothing effect by activating thermosensory pathways associated with comfort, safety, and autonomic regulation, while tactile vibration provides a more immediate sensory input through mechanoreceptor pathways linked to parasympathetic modulation and emotional regulation. However, when used independently alone, thermal stimulation may have a slower perceptual onset, while tactile stimulation may lack sustained affective depth. Combining both modalities may therefore provide a broader, more adaptable sensory experience by integrating the emotional warmth of thermal input with the dynamic regulatory effects of tactile stimulation. In this study, a non-invasive multimodal scalp stimulation device was developed to deliver localized thermal and tactile stimulation inputs while allowing simultaneous electroencephalogram (EEG) recording. The device used Peltier modules for thermal stimulation at 39°C, 40°C, and 41°C, coin-type or linear resonant actuators for tactile stimulation at 20, 30, and 40 Hz, and an ESP32-based control architecture with supporting components, including MOSFETs, resistors, drivers, and temperature-monitoring sensors. Participants undergo baseline EEG recording, cognitive stress induction using

Stroop and arithmetic tasks, post-stress EEG recording, eight-minute stimulation exposure, and post-intervention EEG recording with subjective stress and comfort ratings. EEG responses were evaluated through spectral and nonlinear analyses to determine neurophysiological changes associated with thermal-only, tactile-only, placebo, and combined stimulation conditions. This study positions EEG-guided thermo-tactile stimulation as a personalized, non-invasive, and scalable approach for stress intervention in university populations.

**Keywords:** Multisensory Stimulation, EEG-based Neurofeedback, Thermo-Tactile Integration, Stress Modulation, Personalized Intervention

## References

[1] Brelet, L., & Gaffary, Y. (2022). EEG-based assessment of stress modulation using sensory stimuli. *International Journal of Psychophysiology, 177*, 30 to 37.

[2] Yang, C., et al. (2023). Thermal stimulation in stress modulation: An EEG study. *Biological Psychology, 176*, 108587.

[3] Lopes Alves, L., et al. (2025). Electrical stimulation and vibration therapy for neuromodulation. *Biomedical Engineering Online, 24*(1), Article 12.

[4] Zhou, L., et al. (2023). Combining vibration and warmth for emotion regulation: A pilot study. *IEEE Transactions on Affective Computing*.

[5] Stefaniak, A., et al. (2021). Mapping stress-responsive brain regions: A neuroacupuncture approach. *Brain Sciences, 11*(12), 1612.

# Factors Associated with Dietary Supplement Use Among Filipino National Athletes: A Cross-Sectional Study

**Asst. Prof. Karen Leslie Pineda[1*], MSPH, Catherine Joy Escuadra, PhD[2]**

[1]Department of Sports Science, University of Santo Tomas, Manila, Philippines
[2]Department of Physical Therapy, University of Santo Tomas, Philippines
[1] * *Corresponding author: klpineda@ust.edu.ph*

Dietary supplement use among athletes has become increasingly prevalent across various sports disciplines; however, inconsistencies in definitions and measurement approaches have led to conflicting evidence. Although international studies have examined supplement use, limited evidence exists on its prevalence among Filipino national athletes. Understanding the prevalence and factors associated with dietary supplement use is essential for developing athlete-centered educational strategies and promoting evidence-based practices in sports nutrition. This study aims to (1) determine the prevalence of dietary supplement use among Filipino national athletes and (2) examine its association with demographic and sport-related factors, including age, gender, sports discipline, training experience, education level, sources of information, and reasons for supplement use. A cross-sectional analytical design will be employed among Filipino athletes aged 18 years and above who are registered under their respective national sports associations and are free from chronic illnesses or medical treatment. Quota sampling will be utilized to ensure representation across sports disciplines. Data will be collected using the Questionnaire on the Use of Dietary Supplements adapted from Broniecka et al. (2025). Approval from the Philippine Sports Commission has been sought, and a minimum of 121 participants will be recruited to achieve 80% statistical power. The questionnaire includes supplement use patterns. Data collection may be conducted online or on-site, depending on participant availability and location, with research assistants available to clarify questions. Descriptive and inferential statistics will be analyzed using RStudio and Jamovi to examine associations between supplement use and selected variables. Ethical approval was obtained from the UST Ethics Review Committee. Findings are expected to bridge local knowledge gaps, enhance athlete health and performance, support strategic nutrition education programs, and strengthen institutional support and policy development related to athlete nutrition and supplementation practices in sports. Moreover, this study forms part of Phase One of the Philippine National Anti-Doping Organization's Project AGILA, a World Anti-Doping Agency-compliant digital ecosystem and Electronic Medical Record system designed to protect athletes from accidental doping, promote the safe and evidence-based use of dietary supplements, and support overall athlete health, readiness, and compliance with anti-doping regulations.

**Keywords:** Dietary supplement, National Athletes, Sports Nutrition, Prevalence

## References

[1] Darvishi L, Askari G, Hariri M, Bahreynian M, Ghiasvand R, Ehsani S, et al. The use of nutritional supplements among male collegiate athletes. Int J Prev Med. 2013;4(Suppl 1):S68-72.

[2] Maughan RJ, Burke LM, Dvorak J, Larson-Meyer DE, Peeling P, Phillips SM, et al. IOC consensus statement: dietary supplements and the high-performance athlete. Br J Sports Med. 2018;52(7):439-55. doi: 10.1136/bjsports-2018-099027.

[3] Garthe I, Maughan RJ. Athletes and supplements: prevalence and perspectives. Int J Sport Nutr Exerc Metab. 2018;28(2):126-38. doi: 10.1123/ijsnem.2017-0429.

[4] Knapik JJ, Steelman RA, Hoedebecke SS, Austin KG, Farina EK, Lieberman HR. Prevalence of dietary supplement use by athletes: systematic review and meta-analysis. Sports Med. 2016;46(1):103-23. doi: 10.1007/s40279-015-0387-7.

[5] Broniecka A, Sarachman A, Zagrodna A, Książek A. Dietary supplement use and knowledge among athletes: prevalence, compliance with AIS classification, and awareness of certification programs. J Int Soc Sports Nutr. 2025;22(1):2496450. doi: 10.1080/15502783.2025.2496450.

# Interprofessional collaboration competency development in healthcare students during clinical placements in the time of COVID-19: a mixed methods systematic review

[1]Roi Charles Pineda, [2]Priya Martin,[3]Kimberly Khor, [4]Jocel M. Regino, [5]Lauren Smith, [6]Romeo-Luis F. Ramirez and [7]Michael Palapal Sy

[1]Research Associate, University of Leuven, Belgium
[2]Associate Professor, University of Southern Queensland, Australia
[3]Medical Student, University of Queensland, Australia
[4]Assistant Professor, University of Santo Tomas, Philippines
[5]Research Assistant, University of Southern Queensland, Australia
[6]Instructor, De La Salle Medical and Health Sciences Institute, Philippines
[7]Senior Researcher, Zurich University of Applied Sciences, Switzerland

## ABSTRACT

**Background**. The COVID-19 pandemic severely disrupted traditional clinical education for healthcare students. Consequently, the delivery and prioritization of interprofessional education and collaborative practice (IPECP) training shifted drastically toward remote and virtual platforms. Because of this shift, it remains unclear how these alternative placement models affected students' collaborative practice readiness.

**Objectives**. This review to identify which IPECP competencies developed in pre-qualification healthcare students while on clinical placements during COVID-19 pandemic and to identify the barriers and enablers in facilitating the development of IPECP competencies during the pandemic.

**Methods.** A mixed-methods systematic review was conducted following PRISMA guidelines. Major databases and supplementary sources were searched for primary research published from 2020 onward. Twenty studies met the inclusion criteria. Extracted data were synthesized using a convergent segregated approach and mapped against the Canadian Interprofessional Health Collaborative (CIHC) Competency Framework.

**Results**. Student perceptions indicate that these alternative placements supported the learning of interprofessional collaboration competencies. Outcomes mapped against the updated Canadian Interprofessional Health Collaborative Competency Framework indicate that the most frequently reported interprofessional collaboration competency was team communication and the least reported were collaborative leadership and team differences/disagreements processing. Although gains in interprofessional collaboration competencies were reported across the studies, their methodological shortcomings make

it difficult to determine whether alternative placements (e.g. online and telephone-based) were better or comparable to traditional placements (i.e. with face-to-face interactions), for interprofessional collaboration competency development.

**Conclusion**. Non-traditional and adapted clinical placement models successfully fostered essential IPECP competencies during the pandemic. These findings offer valuable guidance for designing resilient interprofessional education programs capable of withstanding future public health emergencies or localized disruptions.

Keywords: Clinical clerkship; COVID-19 pandemic; healthcare students; interprofessional education; telehealth

| Enablers | Barriers |
|---|---|
| Communication<br>• Effective communication (Lestari et al., 2023)<br>• Virtual platforms allowing and supporting communication and teamwork (Major et al., 2023) | COVID-19 pandemic restrictions<br>• Reduced opportunities for in-person teamwork and challenges in virtual interactions (Major et al., 2023) |
| Role clarity of team members<br>• Clear role definitions within teams (Lestari et al., 2023) | Limited exposure and practical experience<br>• Restrictions on direct patient care and limited team exposure (Castro et al., 2021; Lestari et al., 2023) |
| Curriculum integration and faculty support<br>• Longitudinal curriculum incorporating IPE (Bautista et al., 2020)<br>• Active faculty support and mentorship (Bautista et al., 2020; Leiman et al., 2021) | Technological and logistical challenges<br>• Technology issues, connectivity, and coordination challenges (Castro et al., 2021; Palmer et al., 2023; Pittman et al., 2023; Walker et al., 2022; Winship et al., 2020) |
| Virtual and hybrid formats offer flexibility<br>• Flexibility in delivery formats of learning activities afforded by virtual IPE activities (Major et al., 2023; Mozer et al., 2023; Neumann-Podczaska et al., 2021; Palmer et al., 2023; Winship et al., 2020) | Curriculum and resource constraints<br>• Traditional curricula and resource limitations (Bautista et al., 2020; Major et al., 2023; Mozer et al., 2023; Taylor et al., 2021)<br>• Need for proper training and development of staff involved in delivering IPE (DaSilva et al., 2022; Palmer et al., 2023) |
| Practical experience and feedback<br>• Hands-on clinical experiences and faculty feedback (Bautista et al., 2020; Castro et al., 2021; DaSilva et al., 2022; Mozer et al., 2023; Taylor et al., 2021) | Uncertainty of expectations<br>• Unclear role expectations and insufficient pre-placement frameworks (Kelly et al., 2023) |

**References:**

Bode, S., Dürkop, A., Wilcken, H., Peters, S., & Straub, C. (2021). Interprofessional learning during SARS-CoV-2 (COVID-19) pandemic conditions: The learning project i-recover as a substitute for a rotation on an interprofessional training ward. *GMS Journal for Medical Education*, *38*(1), 1–8. https://dx.doi.org/10.3205/zma001409

Canadian Interprofessional Health Collaborative. (2024, April). *Cihc competency framework for advancing collaboration*. CIHC Competency Framework for Advancing Collaboration

**PRESENTATION TYPE**

/ Oral Presentation

☐ Poster Presentation

**CONFERENCE TOPIC AREA**

☐ Social Inequality & Economic Disparities

/ Public Health & Well-being

☐ Education & Social Change

☐ Environmental & Sustainability Issues

☐ Technology & Society

☐ Conflict Resolution & Human Rights

☐ Governance & Policy Making

☐ Artificial Intelligence

☐ Wireless Technologies

☐ Cyber Security

☐ Sustainable Technology & Emerging Technologies

☐ Other (please specify): _____

# Creating a Community of Practice for Implementation Capacity Building: Komunidad ng Pagsasanay Para sa Pagpapaunlad ng Kapasidad sa Pagpapatupad- The KAPPA Project

**Serra, Mark Angel B.**[1*]; **Dones, Valentin C. III.**[1]; **Kent, Bridie**[2]; **Underwood, Frazer.**[2,*]

[1] Department of Physical Therapy, University of Santo Tomas, Manila, Philippines

[2] University of Plymouth, England, United Kingdom

***** *Corresponding author: mbserra@ust.edu.ph*

The KAPPA Project (Komunidad ng Pagsasanay para sa Pagpapaunlad ng Kapasidad sa Pagpapatupad) is an international collaboration between UK and Philippine institutions to strengthen the capacity of health professionals to implement evidence-based practice in health, care, and public health settings. It responds to the global need for effective support to translate evidence into practice, particularly in low- and middle-income country contexts where health inequities persist. The project aims to build capacity and capability for evidence implementation, co-create an intercontinental Community of Practice (CoP), and support progress toward JBI endorsement and affiliated group status for Philippine partner centers. It seeks to enhance participants' knowledge, confidence, and leadership in evidence-based practice while generating exemplar implementation projects aligned with local clinical priorities. A year-long program, co-led by a steering group, will adapt a tested UK model using the JBI Evidence Implementation Model and PACES. Twenty participants from Philippine institutions will receive online, "bite-sized" education, develop locally governed implementation projects, and engage in an online CoP blending structured teaching, informal networking, and mentorship from UK-based Filipino health professionals. Evaluation will use pre/post Evidence-Based Practice Questionnaire data and qualitative feedback. Expected results include a sustainable CoP, completion of JBI Evidence Implementation Training by a local trainer, exemplar projects in participating centers, and increased self-reported evidence-based practice capability. The project is expected to embed sustainable implementation expertise, deepen institutional partnerships, and

offer a scalable model to advance health equity through more consistent, contextually adapted use of best evidence in clinical practice.

**Keywords:** *Community of Practice, Evidence Implementation, Capacity Building, Health Professionals, JBI Model, Health Equity*

## References

[1] Underwood F, Kent B, Rae P, O'Keeffe K, Shawe J. *The South West Clinical School model: an initiative to transform clinical academic career pathways*. Journal of Research in Nursing. 2023;28(6-7):531-542. doi:10.1177/17449871231209037

# GENERATIVE AI VS STUDENT: A CROSS-SECTIONAL ANALYSIS OF CRITICAL APPRAISALS IN PHYSICAL THERAPY RESEARCH

L.P.D. Tangcuangco*[1,2], C.J.T. Escuadra[1,2], A.S.P. Caballero, E.A.P. Ferrer[1], J.A.D. Manalo[1], S.A.U. Manubay[1], A.G.M. Rabacal[1], Z.R. Teodosio[1]

*Department of Physical Therapy, College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*[1]

*Center for Health Research and Movement Science, College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*[2]

**Corresponding author:* ldtangcuangco@ust.edu.ph

Critical appraisal is essential for evidence-based practice (EBP) in physical therapy (PT) but faces barriers like time constraints and inconsistent evaluations. Generative AI tools, such as ChatGPT, present an innovative solution by offering rapid and structured appraisals. This study assessed the quality of critical appraisals generated by ChatGPT on PT research by comparing them to the rating scores of critical appraisals conducted by undergraduate PT students. A quantitative, analytic, cross-sectional design was utilized to compare retrospective data from anonymized student appraisals and AI-generated outputs. Using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Randomized Controlled Trials, 27 journal articles were purposively sampled. Descriptive statistics and the Mann-Whitney U Test ($p < 0.05$) were employed to analyze appraisal score differences. Results revealed no significant difference in total appraisal scores between AI and student groups ($p=0.48$). The two groups demonstrated high agreement (74.10%–92.60%) on 11 of 13 JBI items, including randomization, blinding, and statistical analysis. However, significant differences emerged in items requiring mechanistic reasoning and contextual interpretation, specifically the concealment of group allocation ($p=0.01$) and similarity of identical treatment procedures ($p=0.01$). For these items, AI responses exhibited higher variability and tended to evaluate items as "unclear" when explicit statements were absent, whereas human raters deduced context more effectively. ChatGPT's critical appraisal capabilities are comparable to those of PT students for straightforward methodological criteria. However, AI exhibits limitations in interpreting implicit details requiring human judgment. A human-AI collaborative approach is recommended for EBP education, utilizing AI as a supportive preliminary scaffold while relying on human oversight for complex methodological evaluations.

**Keywords:** ChatGPT; physical therapy research; critical appraisal; evidence-based practice; artificial intelligence

| Questions | Similar rating (n,%) | Different rating (n,%) | p-value | Rank Biserial Correlation Effect Size |
|---|---|---|---|---|
| 1 - Was true randomization used for assignment of participants to treatment groups? | 23 (85.20) | 4 (14.80) | 0.064 | -0.173 |
| **2 - Was allocation to treatment groups concealed?** | **15 (55.60)** | **12 (44.40)** | **0.010** | **-0.384** |
| 3 - Were treatment groups similar at the baseline? | 26 (96.3) | 1 (3.70) | 0.336 | -0.037 |
| 4 - Were participants blind to treatment assignment? | 21 (77.80) | 6 (22.20) | 0.844 | 0.025 |
| 5 - Were those delivering treatment blind to treatment assignment? | 20 (74.10) | 7 (25.90) | 0.835 | -0.030 |
| 6 - Were outcomes assessors blind to treatment assignment? | 21 (77.80) | 6 (22.20) | 0.072 | -0.239 |
| **7 - Were treatment groups treated identically other than the intervention of interest?** | **18 (66.70)** | **9 (33.30)** | **0.009** | **-0.340** |
| 8 - Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed? | 24 (88.90) | 3 (11.10) | 0.586 | 0.036 |
| 9 - Were participants analyzed in the groups to which they were randomized? | 25 (92.60) | 2 (7.40) | 0.161 | -0.074 |
| 10 - Were outcomes measured in the same way for treatment groups? | 24 (88.90) | 3 (11.10) | 0.571 | 0.037 |
| 11- Were outcomes measured in a reliable way? | 24 (88.90) | 3 (11.10) | 0.089 | 0.108 |
| 12 - Was appropriate statistical analysis used? | 25 (92.60) | 2 (7.40) | 0.331 | -0.071 |
| 13 - Was the trial design appropriate, and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial? | 24 (88.90) | 3 (11.10) | 0.571 | -0.037 |

Table 1: Frequency and proportion of similar and different ratings given by AI and student raters

References:


1. Sackett, D. L. (1997). Evidence-based medicine. Semin Perinatol, 21(1), 3-5. https://doi.org/10.1016/s0146-0005(97)80013-4
2. The Joanna Briggs Institute. (2017). Checklist for Randomized Controlled Trials. The Joanna Briggs Institute.
3. Tarakji, Z., Kanaan, A., Saadi, S., et al. (2024). Concordance between humans and GPT-4 in appraising the methodological quality of case reports and case series using the Murad tool. BMC Med Res Methodol, 24(1), 266. https://doi.org/10.1186/s12874-024-02372-6
4. Hao, J., Yao, Z., Tang, Y., et al. (2025). Artificial Intelligence in Physical Therapy: Evaluating ChatGPT's Role in Clinical Decision Support for Musculoskeletal Care. Ann Biomed Eng, 53, 9-13. https://doi.org/10.1007/s10439-025-03676-4
5. Zhang, Q., & Rapport, M. J. (2025). Purposeful Integration of Artificial Intelligence in Evidence-Based Practice Course for Doctor of Physical Therapy students. International Journal of Allied Health Sciences and Practice, 23(2).

**PRESENTATION TYPE**

☐ Oral Presentation
✔ Poster Presentation

**CONFERENCE TOPIC AREA**

☐ Social Inequality & Economic Disparities
☐ Public Health & Well-being
✔ Education & Social Change
☐ Environmental & Sustainability Issues
☐ Technology & Society
☐ Conflict Resolution & Human Rights
☐ Governance & Policy Making
☐ Artificial Intelligence
☐ Wireless Technologies
☐ Cyber Security
☐ Sustainable Technology & Emerging Technologies
☐ Other (please specify): _____

# Exploring Teachers' Perspectives on Facilitating Social Skills Among Learners with Special Needs in Inclusive Schools in Quezon City: A Qualitative Descriptive Study

**Bautista, Elinor C.; Faminial, Shyruel Rose M.; Cabradilla, Kirsten Carys A.; Mamitag, Shen Mee Coleen L.; Mangahas, Marielle Anne; Sindayen, Patrick D.; Ymballa, Jianna B.**

[1]*Department of Speech-Language Pathology, University of Santo Tomas, Manila, Philippines*
[2]*College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*
** Corresponding author: elcbautista@ust.edu.ph*

Social skills play a vital role in a student's academic performance and overall development. Although speech-language pathologists directly target social skills, teachers are considered critical agents in their ongoing facilitation, as they spend most of their time in the classroom. This study aims to describe the perspectives of teachers on the strategies to support social skills and the factors that hinder their facilitation. A qualitative descriptive design was employed where nine public and private inclusive education teachers were recruited through purposive sampling. Data was collected through face-to-face and online interviews using a semi-structured interview guide. All data were analyzed using Braun and Clarke's six-step thematic analysis framework, with NVivo software used to aid the analysis. Three themes emerged from the study: (1) Teachers' Notions of Social Skills, (2) Teachers Create Interactions in Inclusive Classrooms, and (3) Teachers Recognize the Multifaceted Contextual Factors to Social Inclusion. The findings of the study show that social skills are defined as essential competencies that enable learners to interact and adapt to social contexts, making them crucial for long-term development. Teachers also view themselves as facilitators of social skills through individualized instruction, providing opportunities for peer interaction, fostering peer-mediated support, and interprofessional collaboration. However, barriers such as limited knowledge and training, persistent stigma, and lack of family support hindered effective facilitation. In conclusion, teachers' efforts to facilitate social skills are shaped by contextual factors, often creating a gap between their intentions and capacity. Stronger systemic support, collaboration, and sustained professional development are required for effective facilitation.

**Keywords:** social skills, inclusive schools, learners with special needs

**PRESENTATION TYPE**

☐ Oral Presentation
✔ Poster Presentation

**CONFERENCE TOPIC AREA**

☐ Social Inequality & Economic Disparities
☐ Public Health & Well-being
✔ Education & Social Change
☐ Environmental & Sustainability Issues
☐ Technology & Society
☐ Conflict Resolution & Human Rights
☐ Governance & Policy Making
☐ Artificial Intelligence
☐ Wireless Technologies
☐ Cyber Security
☐ Sustainable Technology & Emerging Technologies
☐ Other (please specify): _____

# Exploring Interventions and Strategies Used by School-Based Speech-Language Pathologists for Grade School Students within the National Capital Region of the Philippines: A Qualitative Descriptive Study

**Blanco, Pia Angela, J.; Bautista, Elinor C.; Agbay, Czarine Aziza F.; Gamo, Marcus Gerome D.; Javier, Jorjana Betina N.; Santiago, Allyssa Fayth R.; Sison, Carl Angela M.; Weir, Krizzia Cybelle P.**

*Department of Speech-Language Pathology, University of Santo Tomas, Manila, Philippines*
*College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*
** Corresponding authors: elcbautista@ust.edu.ph, pjblanco@ust.edu.ph*

Communication disorders may create long-term barriers that can impact children, leading to difficulties catching up academically and socially [1]. In school settings, Speech-language Pathologists (SLPs) play an essential role in supporting students' communication, literacy, and participation through individualized and collaborative interventions [2]. Despite the growing recognition of school-based SLPs in the Philippines, limited local research has explored their intervention practices and decision-making processes [3]. This qualitative descriptive study explored the interventions, strategies, and contextual factors influencing the practices of school-based SLPs working with grade school students within the National Capital Region (NCR) of the Philippines. Ten licensed school-based SLPs with at least two years of experience participated in semi-structured one-on-one interviews conducted through Zoom. Data gathered were transcribed verbatim and analyzed using thematic analysis. Findings revealed that school-based SLPs implement diverse intervention approaches depending on the communication and academic needs of students. Language-focused interventions, literacy-based supports, speech sound motor-based approaches, cognitive and academic supports, behavioral strategies, and sensory-based approaches were commonly utilized by school-based SLPs working in the NCR. These are delivered through push-in and pull-out models, individual, dyad, and group-based sessions, which depend on students' readiness, classroom demands, and institutional factors. These choices were shaped by theoretical knowledge, ecologically valid assessment practices, collaboration with teachers and other professionals, and the individual strengths and needs of students. However, contextual constraints such as scheduling conflicts, large caseloads, differing professional perspectives, and limited institutional support affected intervention implementation. Overall, the findings highlight how school-based SLPs in the NCR adopt flexible, holistic, and context-responsive intervention practices to support students' communication and academic participation. The study underscores the need for localized research, stronger institutional support, and more contextually responsive school-based speech-language pathology practices in the Philippines.

**Keywords:** school-based SLPs, strategies, interventions, Qualitative Descriptive, NCR

## References

[1] Riad, R., Allodi, M. W., Siljehag, E., & Bölte, S. (2023). Language skills and well-being in early childhood education. *Frontiers in Education, 8*, 963180.

[2] Steele, S. C. (2020). Vocabulary intervention: A national survey. *Communication Disorders Quarterly, 41*(3), 151–161.

[3] Denman, D., Cordier, R., Kim, J. H., Munro, N., & Speyer, R. (2021). What influences speech-language pathologists' use of different types of language assessments for elementary school-age children? *Language, Speech, and Hearing Services in Schools, 52*(3), 776–793.

# Perspectives of Occupational Therapy Clinicians on the Use of Artificial Intelligence in Clinical Practice in Metro Manila: A Qualitative Study

**Gomez, Ivan Neil B.[1]; Cua, Justine Anne G.[1]; Endozo, Catherina Moira A.[1,*]; Berja, Kariza Gale F.[1]; Go, Gabriel Derick C.[1]; Lagman, Sabina Diorela Simone G.[1]; Rodriguez, Jenny Lynn S.[1]**

[1] Department of Occupational Therapy, University of Santo Tomas, Manila, Philippines

***** *Corresponding author: catherinamoira.endozo.crs@ust.edu.ph*

Artificial intelligence (AI) is increasingly transforming healthcare, including its emerging applications in occupational therapy (OT), with generative artificial intelligence (GenAI) showing potential to enhance clinical practice [1]-[3]. However, its integration depends not only on technological advancements but also on clinicians' perceptions, acceptance, and experiences, which remain underexplored [3]. This qualitative phenomenological study aimed to explore the perspectives of OT clinicians in Metro Manila, Philippines on the use of AI, particularly GenAI, in clinical practice, focusing on their experiences, attitudes, acceptance, and intention to use the technology. Guided by the Technology Acceptance Model (TAM), the study involved 16 OT clinicians working across Metro Manila who were familiar with AI and with at least one year of clinical experience, excluding those in academic roles. Participants were also purposively sampled to ensure variation in fields of practice and clinical settings. Following a successful pilot test, one-on-one semi-structured interviews were conducted online and thematically analyzed using NVivo 15 based on a coding framework. Five themes emerged: (i) AI as a supportive but bounded clinical tool, (ii) safeguarding professional identity and clinical judgment, (iii) differences in user competence shaping effective AI use, (iv) conditional acceptance of AI based on perceived risks, and (v) context-informed recommendations for AI use and improvement. Participants generally viewed AI as useful for improving efficiency in clinical and administrative tasks, but expressed concerns regarding reliability, ethics, privacy, and contextual resource constraints. Overall, acceptance of AI was generally positive but cautious, with its use seen as highly context-dependent and requiring professional judgment. The findings highlight the need for training, clear guidelines, and responsible integration to support effective and ethical use in OT practice.

**Keywords:** artificial intelligence; occupational therapy; perspectives; perceptions; qualitative

**Table 1.** Summary of Themes and Sub-themes Generated in this Study

| Theme | Sub-Themes |
|---|---|
| (1) AI as a Supportive but Bounded Clinical Tool | 1.1 AI as a supportive adjunct tool in practice<br>1.2 AI as an idea generator and efficiency enhancer |
| (2) Safeguarding Professional Identity and Clinical Judgment | 2.1 Preserving clinical judgment in AI-assisted practice<br>2.2. Concerns with the reliability and regulation of AI |
| (3) Differences in User Competence Shape Effective AI Use | 3.1 Varying knowledge and skills with emerging AI<br>3.2 Understanding AI as an adaptive yet user-dependent tool |
| (4) Conditional Acceptance of AI Based on Perceived Risks | 4.1 Receptive yet guarded views on AI use<br>4.2 Resistance shaped by perceived risks of AI |
| (5) Context-Informed Recommendations for AI Use and Improvement | 5.1 Contextual opportunities and constraints in AI use<br>5.2 Advancing systems and capacities for responsible AI use |

## References

[1] Wang, F. & Preininger, A. (2019). AI in Health: State of the Art, Challenges, and Future Directions. *Yearbook of Medical Informatics*, 28(1), 16–26. https://doi.org/10.1055/s-0039-1677908.

[2] Hirani, R., Noruzi, K., Khuram, H., Hussaini, A., Aifuwa, E.I., Ely, K., et al., (2024). Artificial Intelligence and Healthcare: A Journey through History, Present Innovations, and Future Possibilities. Life 14(5), 557. https://doi.org/10.3390/life14050557.

[3] Nah, F.F.-H., Zheng, R., Cai, J., Siau K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, 25(3), 277-304. https://doi.org/10.1080/15228053.2023.2233814

# Outcomes Insights: A Learning Analytics Dashboard for Entry-Level Physical Therapy Learning Outcomes in a University in the Philippines

C.J.T. Escuadra[1,2]

[1]Department of Physical Therapy, College of Rehabilitation Sciences, University of Santo Tomas
[2]Department of Education, Ewha Womans University
*Corresponding author: ctescuadra@ust.edu.ph*

The use of a learning analytics dashboard (LAD) is a systematic way to analyze and monitor program-intended learning outcomes and to provide feedback to help improve students' learning strategies. Giving students the opportunity to view and understand their status in such an application allows them to be more involved and empowered in their learning process. Past studies in Namoun and Alshanqiti's (2020) review have highlighted that each institution should create its own learning analytics system to account for the specific characteristics of its students and curriculum. This study aims to develop and test a prototype LAD called Outcomes Insight and to determine its effectiveness in the learning and study strategies of selected Filipino university students. This research study had four sequential phases following an iterative design-based research described by Wang and Hannafin (2015) and Yoo & Jin (2020). In phase 1, predictive models were developed using structural equation modeling and artificial neural network (SEM-ANN) on the final grades of 399 physical therapy students. In phase 2, Outcomes Insight was developed using the predictive models and evaluated by three education experts. In phases 3-4, the acceptability and effects of the dashboard on the learning and study strategies of 28 second-year physical therapy students were tested using the Learning and Study Strategies Inventory (LASSI) 3rd edition. Data from phases 3 and 4 were analyzed using descriptive and inferential statistics. For phase 1, relevant courses for predicting each program outcome were identified with acceptable fit metrics in SEM-ANN (CFI & TLI>0.90, RMSEA & SRMR<0.08, MAE & MSE=0.071-0.418). For phase 2, the developed prototype included three distinct but related pages: input, results, and feedback pages. Experts and students agreed that the dashboard includes relevant elements and a visually appealing, accessible, and easy-to-use structure. For Phase 4, significant changes ($p<0.05$) and medium to large effect sizes (0.520-0.881) were found across students' habits, professional development, and mind. The developed and validated Outcomes Insight prototype presents an excellent potential for providing descriptive, predictive, and prescriptive feedback to improve physical therapy students' learning and study strategies. The dashboard included comprehensive, accurate predictive models and personalized recommendations to help students understand their current status and potential improvement options. Health educators should continuously use innovative strategies and relevant data to monitor students' progress and provide personalized recommendations.

Keywords: Learning analytics dashboard, Structural Equation Modeling, Artificial Neural Network

**Table 1.** *Summary of LASSI Scores of students at pre and post-tests*

| (Total=28) | *Pre-Test* | | *Post-Test* | | *Statistic* | *p-value* | *Effect Size* | Interpretation |
|---|---|---|---|---|---|---|---|---|
| | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Habits of Professionals | | | | | | | | |
| Attitude | 43.10 | 23.50 | 54.00 | 23.90 | 78.0 | 0.023 | 0.520 | Moderate |
| Motivation | 44.30 | 27.10 | 52.80 | 26.40 | 84.0 | 0.020 | 0.521 | Moderate |
| Time Management | 49.10 | 24.0 | 50.60 | 23.10 | 167.5 | 0.848 | 0.046 | - |
| Average | 44.90 | 20.50 | 52.50 | 18.80 | 64.5 | 0.013 | 0.659 | Large |
| Habits of Learning | | | | | | | | |
| Information Processing | 74.40 | 19.50 | 74.30 | 21.50 | 106.0 | 0.672 | 0.116 | - |
| Selecting Main Ideas | 37.10 | 25.90 | 47.00 | 26.90 | 73.0 | 0.049 | 0.470 | Moderate |
| Self-Testing | 75.30 | 19.90 | 81.20 | 13.40 | 80.0 | 0.075 | 0.420 | - |
| Test Strategies | 40.00 | 23.50 | 45.20 | 24.20 | 118.5 | 0.239 | 0.271 | - |
| Using Academic Resources | 35.50 | 24.80 | 51.60 | 23.40 | 12.5 | <0.001 | 0.881 | Large |
| Average | 52.40 | 15.00 | 59.90 | 12.60 | 41.0 | <0.001 | 0.766 | Large |
| Habits of Mind | | | | | | | | |
| Anxiety | 20.60 | 19.90 | 35.50 | 23.50 | 34.5 | 0.002 | 0.750 | Large |
| Concentration | 36.60 | 28.30 | 44.00 | 24.00 | 50.5 | 0.075 | 0.468 | - |
| Average | 28.50 | 19.50 | 39.70 | 19.50 | 33.5 | <0.001 | 0.809 | Large |

## References

[1] Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), 237.

[2] Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. Educational technology research and development, 53(4), 5-23.

[3] Yoo, M., & Jin, S. H. (2020). Development and evaluation of learning analytics dashboards to support online discussion activities. Educational Technology & Society, 23(2), 1-18.

**PRESENTATION TYPE**

☐ Oral Presentation

■ Poster Presentation

**CONFERENCE TOPIC AREA**

☐ Social Inequality & Economic Disparities
☐ Public Health & Well-being
☐ Education & Social Change
☐ Environmental & Sustainability Issues
☐ Technology & Society
☐ Conflict Resolution & Human Rights
☐ Governance & Policy Making
☐ Artificial Intelligence
☐ Wireless Technologies
☐ Cyber Security
■ Sustainable Technology & Emerging Technologies
☐ Other (please specify): _____

# Characterizing Musical Parameters for Stress Intervention: A Biofeedback-Based Analysis of Physiological and Neurophysiological Responses

**Fabay, Lianna Mae A.**[1,2,*]; **Ocenar, Samuel John A.**[1,2]; **Danganan, Lester Mico L.**[5]; **Nieles, Jay Patrick M.**[1]; **Catindig-Dela Rosa, Joseree Ann S.**[6,7]; **Prado, Seigfred V.**[1,2,3,4]

[1] Brain Research Advancement and Innovations in Neurotechnology (BRAIN) Lab, University of Santo Tomas, Manila, Philippines
[2] Graduate School, University of Santo Tomas, Manila, Philippines
[3] Department of Electronics Engineering, University of Santo Tomas, Manila, Philippines
[4] Research Center for Engineering and Technological Sciences, University of Santo Tomas, Manila, Philippines
[5] NSW Health, Sydney, New South Wales, Australia
[6] Faculty of Medicine and Surgery, University of Santo Tomas, Manila, Philippines
[7] National Center for Geriatric Health, Manila, Philippines
* *Corresponding author: liannamae.fabay.gs@ust.edu.ph*

Stress is a psycho-physiological response to overwhelming demands, often triggered by academic pressures, and is most prevalent among college students, reporting the highest levels of perceived stress. Pharmacological coping strategies, such as the use of stimulants and sedatives, are widely practiced but pose risks of dependency, cognitive fatigue, and limited long-term effectiveness. As a result, safer and more sustainable alternatives are necessary. Among the available non-pharmacological approaches, music therapy (MT) has consistently demonstrated superior efficacy in reducing stress and anxiety due to its influence on both emotional regulation and autonomic control. While MT has shown promising results in improving mood and lowering physiological arousal, most current studies rely on subjective reporting and lack objective neurophysiological evaluation. This study developed a quantitative framework to evaluate the effects of musical parameter combinations on EEG-based neural responses and on smartwatch-monitored physiological responses during a music-based stress intervention. Musical stimuli were selected from the DEAM dataset using a feature-based modeling approach, in which extracted audio features were reduced via Principal Component Analysis and grouped using *K*-means clustering, yielding 12 representative music excerpts. Neurophysiological data were recorded using an 8-channel OpenBCI EEG system, while physiological markers, including heart rate, electrodermal activity, and skin temperature, were captured using an EmotiBit physiological monitoring smartwatch. Lab Streaming Layer (LSL) was used to synchronize multimodal signals across baseline, stress, music listening, and post-intervention

phases. EEG and physiological data were processed to extract stress-relevant features and were prepared for analysis using Canonical Correlation Analysis (CCA) and Linear Mixed-Effects Modeling (LMM). Through this integrative approach, the study provided an objective basis for identifying music parameters that support personalized, scalable, and non-invasive stress intervention.

**Keywords:** Music Therapy, Stress Intervention, EEG, Physiological Monitoring, Multimodal signals

## References

[1] Barbayannis G, Bandari M, Zheng X, Baquerizo H, Pecor KW, Ming X. Academic Stress and Mental Well Being in College Students: Correlations, Affected Groups, and COVID 19. *Frontiers in Psychology*. 2022;13:886344. doi: 10.3389/fpsyg.2022.886344.

[2] Fedotchev, A., Parin, S., Polevaya, S., & Zemlianaia, A. (2022). EEG based musical neurointerfaces in the correction of stress induced states. *Brain Computer Interfaces, 9*(1), 1 to 6.

[3] Amanvermez, Y., Rahmadiana, M., Karyotaki, E., De Wit, L., Ebert, D. D., Kessler, R. C., & Cuijpers, P. (2020). Stress management interventions for college students: A systematic review and meta analysis. *Clinical Psychology Science and Practice, 30*(4), 423 to 444. https://doi.org/10.1111/cpsp.12342

[4] Paszkiel S, Dobrakowski P, Łysiak A. The Impact of Different Sounds on Stress Level in the Context of EEG, Cardiac Measures and Subjective Stress Level: A Pilot Study. *Brain Sciences*. 2020;10(10):728. doi: 10.3390/brainsci10100728.

[5] El Makrini, I., Turcksin, T., Incirci, T., Thiery, E., Kindt, S., Lovecchio, R., Cao, H. L., Denayer, M., Lamine, E., Huysentruyt, S., Verstraten, T., & Vanderborght, B. (in press). PhysioSense: An Open Source Multi Modal Monitoring Framework for Human Movement and Behavior Analysis. *IEEE Robotics & Automation Magazine*.

# Perspectives and Experiences of Physical Therapy Interns on Telerehabilitation in the Philippines

**Fernandez, Ma. Roxanne L.[1],[2]*; Cruz, Christopher G.[1]; Co, Mary Therese A.[1]; Aypa, Nicolas Macwain G.[1]; Dy-Lim, Vanna Chelsey C.[1]; Molina, Philip Louie L.[1]; Pasco, Julianne Kimberly U.[1]; Relucio, Russ Jason L.[1]; Ugbinar, Asia C.[1]**

[1] Department of Physical Therapy, University of Santo Tomas, Manila, Philippines
[2] Center for Health Research and Movement Science, University of Santo Tomas, Manila, Philippines
* *Corresponding author: mlfernandez@ust.edu.ph*

The COVID-19 pandemic accelerated the integration of telerehabilitation into physical therapy (PT) education and practice, aiming to maintain accessibility, affordability, and safety in patient care. By using tools such as video conferencing and remote exercise monitoring, telerehabilitation supported clinical learning during rotations. Understanding the perspectives and experiences of PT interns offers important insights into their role in patient care and future curriculum development. This study explored the perspectives and experiences of PT interns regarding telerehabilitation, focusing on preparedness, challenges, adaptive strategies, and perceptions of clinical effectiveness. A descriptive qualitative design guided by the Normalisation Process Theory (NPT) and the Unified Theory of Acceptance and Use of Technology (UTAUT). Participants were recruited through purposive sampling. While recruitment targeted PT graduates nationwide who had prior telerehab experience, participants who met the eligibility criteria came from Luzon and Iloilo in the Visayas region. Three online focus group discussions were conducted using a researcher-developed interview guide. Sessions were recorded, transcribed verbatim, and analysed using Braun and Clarke's six-step thematic analysis. Thematic analysis generated several key themes: Adapting to a New Mode of Practice, Logistical Barriers, Documentation and Assessment Demands, Home-Based Relevance, Insufficient Preparation, Competence through Experience, Supervisory Support, Technological Familiarity, Limitations of Remote Assessments, Developing Professional Composure, Confronting Contextual and Technical Barriers, Increased Reach, Early Doubts, Growing Engagement, Influence of Caregiver Support, Perceived Effectiveness of Telerehabilitation, Impact of Realistic Goal-Setting, Evaluation of Technologies and Tools Used, Technological Barriers, Sustainability and Accessibility, Educational Integration, and Professional Disposition. In conclusion, telerehabilitation was perceived as a meaningful yet demanding form of care, highlighting the need for structured training, clear protocols, and adequate supervision to support its effective integration into PT education and practice.

**Keywords:** telerehabilitation; physical therapy education; clinical education; qualitative research; normalisation process theory; unified theory of acceptance and use of technology

# Feature Characterization using IMU Kinematic Data to Aid in Diagnosing Patellofemoral Pain Syndrome

***Xhan Benedict Grefaldeo[1], Jade Marco L. Idio[1], Josiah David L. La Guardia[1], Caesar Manuel S. Marquez[1], Consuelo B. Gonzales-Suarez[2,3], Ivan Neil B. Gomez[4], Jan-Tyrone C. Cabrera[2,3], Jazzmine Gail S. Flores[3], Edison A. Roxas[1], Mary Grace Ann C. Bautista[1], Emmanuel C. Guevara[1], Gabriel Rodnei M. Geslani[1]**

*[1] Department of Electronics Engineering, University of Santo Tomas, Manila, Philippines*
*[2] Department of Physical Medicine and Rehabilitation, University of Santo Tomas, Manila, Philippines*
*[3] Research Center for the Health Sciences, University of Santo Tomas, Manila, Philippines*
*[4] Department of Occupational Therapy, University of Santo Tomas, Manila, Philippines*
**Corresponding author: gmgeslani@ust.edu.ph*

This study addresses the diagnostic limitations of binary Patellofemoral Pain Syndrome (PFPS) classification by utilizing Inertial Measurement Unit (IMU) data to characterize PFPS-indicative kinematic features. Principal Component Analysis (PCA) was applied to high-dimensional movement data, retaining 16 components that explained 90.17% of the total variance. These components were evaluated using filter, embedded, and wrapper feature selection methods to identify those most relevant for PFPS classification. Preliminary findings show that the wrapper method achieved the highest balanced accuracy of 76% using two principal components (PC9 and PC16), while all selection frameworks consistently converged on two biomechanical domains: hip-femur rotational mechanics and lumbar trunk coordination. Key features included internal hip and femur rotation and increased lumbar rotational excursion. These results provide proof-of-concept evidence that PFPS is better characterized by coordinated multi-segment interactions rather than isolated joint abnormalities. The findings suggest strong generalization potential for identifying biomechanical signatures of PFPS, supporting future work toward improved clinical diagnosis and personalized treatment strategies.

**MASP 2026** | Abstract Submission June 24-28, 2026 | Manila

# Satisfaction of UST-CRS Students with the Quality of Blended Teaching and Learning in the Interprofessional Education Course for Academic Year 2025-2026: A Cross-Sectional Study

**Dones, Valentin C. III[1]; Versales, Cristine Rose S.[1]; Alvaro, Donita Faye D.[1]; Carpio, John Daniel D.[1]; Garces, Maiquela V.[1]; Ledesma, Antonius Andrei L.[1*]; Sison, Juan Antonio M.[1]; Tabios, Kate Riane S.[1]**

[1] Department of Physical Therapy, University of Santo Tomas, Manila, Philippines

***** *Corresponding author: andreiledesma2@gmail.com*

Blended teaching and learning (BTL) has become increasingly integrated into higher education settings as it combines face-to-face and online learning modalities in order to ensure greater flexibility, engagement, and accessibility. In the Philippines, BTL is aligned with the Commission on Higher Education guidelines. However, local evidence regarding its implementation in health-allied education remains limited. This study evaluated the quality of BTL implementation in the Interprofessional Education (IPE) course of the University of Santo Tomas College of Rehabilitation Sciences (UST-CRS) for Academic Year 2025-2026 by assessing the satisfaction of third-year rehabilitation science students. For this study, a descriptive observational cross-sectional design was used and all eligible third-year UST-CRS students enrolled in the IPE course were invited to participate. Data was collected online via Google Forms using the Blended Teaching Assessment Tool (BTAT), a 37-item questionnaire that measures satisfaction across six dimensions: Course Direction and Guide, Quality Teaching Delivery Skills, Quality of Instructional Media and Recorded Videos, Use of Educational Technology and Instructional Media, Quality of Learning Activities, and Reflective Feedback. A total of 102 students completed the questionnaire, representing a 44% response and completion rate. Descriptive statistics, including mean scores, standard deviations, and 95% confidence intervals, were used to analyze the data gathered. Results showed generally favorable satisfaction across all six BTAT dimensions, with mean scores ranging from 3.075 to 3.561. Dimension III, Quality of Instructional Media and Recorded Videos, received the highest mean score, suggesting strong satisfaction with the accessibility and quality of learning materials. Conversely, Dimension VI, Reflective Feedback, obtained the lowest mean score and highest variability, indicating inconsistencies in the frequency and quality of feedback provided to students. The findings suggest that BTL implementation in the IPE course is generally satisfactory and effective in supporting student learning. However, improvements in standardized and timely reflective feedback may further enhance course delivery, student motivation, and curriculum quality in health-allied education.

**Keywords:** blended learning; student satisfaction; health-allied education; Philippines

**PRESENTATION TYPE**

☐ Oral Presentation

√ Poster Presentation

**CONFERENCE TOPIC AREA**

☐ Social Inequality & Economic Disparities

☐ Public Health & Well-being

☐ Education & Social Change

☐ Environmental & Sustainability Issues

☐ Technology & Society

☐ Conflict Resolution & Human Rights

☐ Governance & Policy Making

☐ Artificial Intelligence

☐ Wireless Technologies

☐ Cyber Security

☐ Sustainable Technology & Emerging Technologies

√ Other (please specify): Sports Psychology and Athletic Development

# Psychological Profiling of Mental Toughness, Motivation, and Grit in UST Men's and Women's Collegiate Basketball Players During the Off-Season: A Descriptive Correlational Study

R.D. Ramos[1], L.R. Cosep[1], E.N. De Mesa[1], J. Dogillo[1], A. See[1], K. Archog[1], E.Coronel[1], J. Lazo[1], H. Perez[1]

[1] *Department of Sports Science, University of Santo Tomas, Manila, Philippines*

**Corresponding author: rdramos@ust.edu.ph*

Psychological attributes such as mental toughness, motivation, and grit are fundamental in shaping how athletes respond to training demands, cope with pressure, and sustain long-term performance. While these constructs have been extensively examined during highly competitive in-season periods, limited research has explored their expression during the off-season – a critical developmental phase characterized by reduced external reinforcement, lower immediate pressure, and a greater reliance on self-regulation and internal drive [1]. Furthermore, profiling these specific psychological traits among Southeast Asian collegiate populations remains underexplored. To address this gap, this study aimed to systematically describe the psychological profile of the University of Santo Tomas (UST) Men's and Women's Collegiate Basketball Teams during the off-season and to examine the underlying relationships among mental toughness, sport motivation, and grit. A descriptive correlational cross-sectional research design was employed, utilizing total enumeration sampling to recruit 22 eligible collegiate basketball players. Participants completed three validated self-report measures: the Sports Mental Toughness Questionnaire (SMTQ-14) [2], the Sport Motivation Scale-II (SMS-II) [3], and the Domain-Specific Grit Scale for College Athletic Students (DGSCAS) [4]. Descriptive statistics were used as a primary analytical tool to summarize and quantify the baseline psychological profiles of the sample, providing a detailed overview of the athletes' psychological variables across specific subscales. The descriptive analysis revealed that athletes maintained moderate to high levels across all main constructs. Specifically, mental toughness ($M = 2.96 \pm 0.38$) showed higher mean scores in constancy ($M = 3.32 \pm 0.57$) and confidence ($M = 3.20 \pm 0.54$) than in emotional control ($M = 2.23 \pm 0.56$), suggesting a potential area for off-season psychological development. Motivation ($M = 5.33 \pm 0.74$) was predominantly driven by self-determined forms, including identified ($M = 6.06 \pm 1.36$), integrated ($M = 5.89 \pm 1.38$), and intrinsic regulation ($M = 5.82 \pm 1.54$). Grit ($M = 4.00 \pm 0.75$)

demonstrated strong perseverance of effort (M = 4.10 ± 0.82) and consistency (M = 3.90 ± 0.82) despite the lack of competitive stimuli. Pearson correlation analysis revealed that mental toughness was significantly and positively correlated with both motivation (r = 0.483, $p$ = 0.023) and grit (r = 0.579, $p$ = 0.005), stating that athletes with higher levels of mental toughness also tend to exhibit higher levels of motivation and grit. Conversely, no statistically significant relationship was observed between motivation and grit (r = 0.282, $p$ = 0.204), stating that these constructs operate independently in this context and that the presence of one attribute does not necessarily correspond to higher levels of the other. The findings suggest that mental toughness serves as a central psychological anchor associated with both motivational processes and sustained effort. These baseline descriptive and correlational insights can critically inform coaches and sports psychologists in designing targeted mental conditioning strategies, particularly interventions aimed at enhancing emotional control and sustaining intrinsic drive outside of the competitive season.

**Keywords:** mental toughness, sport motivation, grit, off-season, collegiate basketball

**Table 1.** Descriptive Statistics and Normality of Study Variables (N = 22) & Relationship Among Main Constructs

| Variables | Level | Variable | Post-Intervention (M ± SD) | Variables | r | *p-value* |
|---|---|---|---|---|---|---|
| Mental Toughness | Overall | SMTQ Total | 2.96 ± 0.38 | Mental Toughness to Motivation | .483 | 0.023 |
| | Subscale | Confidence | 3.20 ± 0.54 | | | |
| | | Constancy | 3.32 ± 0.57 | | | |
| | | Control | 2.23 ± 0.56 | | | |
| Motivation | Overall | SMS Total | 5.33 ± 0.74 | Mental Toughness to Grit | .579 | .005 |
| | Subscale | Intrinsic | 5.82 ± 1.54 | | | |
| | | Integrated | 5.89 ± 1.38 | | | |
| | | Identified | 6.06 ± 1.36 | | | |
| | | Amotivation | 5.06 ± 1.40 | | | |
| Grit | Overall | DGSCAS Total | 4.00 ± 0.75 | Motivation to Grit | .282 | .204 |
| | Subscale | Consistency | 3.90 ± 0.82 | | | |
| | | Perseverance | 4.10 ± 0.82 | | | |

## References

[1] Chou, C., Wang, F., Wu, H., Tsai, S., Chen, C., Bernard, J. R., Kuo, Y., & Liao, Y. (2021). "The Competitive Season and Off-Season": Preliminary Research concerning the Sport-Specific Performance, Stress, and Sleep in Elite Male Adolescent Basketball Athletes. *International Journal of Environmental Research and Public Health, 18*(24), 13259.

[2] Sheard, M., Golby, J., & Van Wersch, A. (2009). Progress toward Construct Validation of the Sports Mental Toughness Questionnaire (SMTQ). *European Journal of Psychological Assessment, 25*(3), 186-193.

[3] Pelletier, L. G., Rocchi, M. A., Vallerand, R. J., Deci, E. L., & Ryan, R. M. (2012). Validation of the revised sport motivation scale (SMS-II). *Psychology of Sport and Exercise, 14*(3), 329-341.

[4] Gao, F., Wei, Q., Dong, X., Gao, J., Lu, S., & Liu, Y. (2024). Developing and validating a Domain-specific Grit Scale for College Athletic Students. *Scientific Reports, 14*(1), 11897.

# Experiences of rehabilitation professionals in palliative care: a qualitative systematic review

**Nikka Karla R. Santos[1,2,*] Jon Timothy M. Rivero[1,3,4] Valentin C. Dones III[1,3]**

[1]The Center for Health Research and Movement Science: A JBI Affiliated Group, University of Santo Tomas, Manila, Metro Manila, Philippines
[2]Department of Occupational Therapy, College of Rehabilitation Sciences, University of Santo Tomas, Manila, Metro Manila, Philippines
[3]Department of Physical Therapy, College of Rehabilitation Sciences, University of Santo Tomas, Manila, Metro Manila, Philippines
[4]The Ruth Foundation for Palliative and Hospice Care, Inc., Muntinlupa, Metro Manila, Philippines
*****Corresponding author: nrsantos@ust.edu.ph*

Rehabilitation professionals play essential but underrecognized roles in palliative care, navigating contrasts between restorative and palliative philosophies, while adapting their practice to prioritise patient comfort, participation, and dignity. Despite growing global recognition of rehabilitative palliative care, misconceptions, systemic barriers, and limited understanding of rehabilitation roles continue to restrict their involvement in palliative care delivery. Qualitative studies exploring the experiences and perspectives of rehabilitation professionals, including occupational therapists, physical therapists, and speech-language pathologists, working in palliative care settings were eligible. Studies using any qualitative design conducted in hospitals, hospices, home-based care, nursing homes, or community-based programs were included. Mixed-methods studies were considered when qualitative findings could be extracted. The review followed JBI methodology for systematic reviews of qualitative evidence. A comprehensive 3-step search strategy was conducted beginning February 20, 2025, across PubMed, CINAHL (EBSCOhost), Embase (Ovid), Web of Science, PEDro, and grey literature sources including ProQuest Dissertations & Theses, WHO resources, and Google Scholar. Search terms for palliative care were adapted from a validated search filter. Two reviewers independently screened titles, abstracts, and full texts, and appraised methodological quality using the JBI Critical Appraisal Checklist for Qualitative Research, with disagreements resolved through discussion or consultation with a third reviewer. Data were extracted using the JBI SUMARI tool and synthesized through JBI's meta-aggregation approach. Confidence in the synthesized findings was assessed using the ConQual approach based on dependability and credibility. The search and citation tracing identified 23,787 citations, of which 19,180 unique records were screened. One hundred eighty-seven reports were assessed, and studies from 16 countries met the inclusion criteria. A total of 277 findings were extracted, of which 275 were included in the synthesis, resulting in 42 categories and eight synthesized findings. These synthesized findings revealed that occupational therapists, physical therapists, and speech-language pathologists use flexible, integrative approaches that support autonomy, meaningful activity, and shared decision making with patients and families. Their interventions address physical, psychosocial, emotional, and spiritual dimensions of care, yet face systemic, cultural, and physician-centred barriers that limit their full integration into interdisciplinary teams. Professionals also reported emotional and relational demands as both burdens and sources of meaning, emphasizing the need to develop coping strategies and reflective practice. Advocacy, training, and policy reform must strengthen the recognition of the rehabilitation professionals' role, expand workforce competence, and embed rehabilitation within national palliative frameworks. Lastly, building compassionate communities through education and partnerships can normalise conversations about death and ensure sustainable, person-centred care beyond institutional settings. The overall ConQual scores for the synthesized findings were rated as low to moderate. Rehabilitation professionals contribute substantially to person-centred palliative care yet remain limited by structural, cultural, and organisational barriers. Strengthening their integration requires targeted education and training, advocacy, and policy reform that embed rehabilitation into national palliative care frameworks. The findings highlight the need to build workforce competence, enhance interdisciplinary collaboration, create standardized practice guidelines, and ensure timely and equitable referral pathways. Future research should further examine culturally responsive approaches to rehabilitative palliative care, assess the effectiveness of rehabilitation interventions across diverse settings, and identify strategies to improve workforce well-being and clarify professional roles.

**Keywords:** person-centred, interdisciplinary, hospice, community-based, occupation

# Design and Development of a Stress Detection System using an Intelligent Pen for the Student's Well-Being

**H.A. Kwan, J.P. Manalo, J.M. Santos, E. Torres, A.B. Valle, E. Roxas**

[1] *Department of Electronics Engineering, University of Santo Tomas, Manila, Philippines*
**Corresponding author: earoxas@ust.edu.ph*

Stress remains a prevalent mental health concern in modern society, necessitating innovative approaches for early detection and intervention. This study proposes the use of iSULAT (Intelligent Stroke Utilization, Learning, Assessment, and Testing), an intelligent pen system utilized to detect mental stress among engineering students at the University of Santo Tomas–Manila through handwriting analysis. The system integrates multiple instruments, including handwriting tasks, the Depression, Anxiety, and Stress Scale-21 (DASS-21), the Visual Analog Scale (VAS), and physiological data from a smartwatch, to provide a comprehensive assessment of stress. Handwriting features captured by the intelligent pen were correlated with physiological indicators, specifically Heart Rate Variability (HRV), and self-reported stress levels from the DASS-21 subscale, both of which correlated with the aspect of mental stress. A data collection of 150 participants was conducted in coordination with the UST Faculty of Engineering, scoping fourth-year engineering students across all the (6) engineering departments, wherein the entire dataset will be analyzed in its complete version and its aligned version. Each configuration was evaluated separately under the machine learning pipeline based on the three chosen algorithms: Support Vector Machine, Decision Tree, and Random Forest to assess how label consistency affects the model's performance. Throughout the entire procedure, it has been detected that air time was the most significant handwriting feature in predicting mental stress, particularly in tasks involving writing the words in capital letters. Among the evaluated models, the Random Forest (RF) algorithm demonstrated an accuracy percentage of 90.91% when identifying the best feature in detecting stress. Additionally, when checking the overall accuracy among all the proposed handwriting tasks, the Decision Tree (DT) algorithm obtained a balanced accuracy of 58.6%, supporting the feasibility of using an intelligent pen system for stress detection within the specified population.

# Perceptions of Generative AI Integration in Clinical Training Among Occupational Therapy Interns in Manila: A Qualitative Study

**Santos, Nikka Karla R.[1,2]*; Fariñas, Maria Ruby P.[1]; Behan, Sean James Eire A.[1]; Agraviador, Ryza Mikyla V.[1]; Buquing, Eladia Denise A.[1]; Diosanta, Josiah Myron A.[1]; Evangelio, Kristin Chloe L.[1]; Santiago, Ma. Dulce Regina D.[1]**

[1] Department of Occupational Therapy, University of Santo Tomas, Manila, Philippines
[2] Center for Health Research and Movement Science, University of Santo Tomas, Manila, Philippines
* *Corresponding author: nrsantos@ust.edu.ph*

Generative Artificial Intelligence (GenAI) is increasingly being integrated in healthcare and rehabilitation education, with potential to support clinical decision-making, documentation, intervention planning, and student learning [1]. Its relevance is expanding in occupational therapy (OT), yet evidence on how GenAI is perceived and used in Philippine OT clinical training remains limited [2, 3]. This study examined the perceptions of OT interns in Manila, Philippines, regarding the integration of GenAI into clinical training, including perceived benefits, challenges, ethical considerations, and preferred approaches to integration, guided by the Diffusion of Innovations Theory. An exploratory qualitative design was employed. Twenty-eight (28) OT interns from two universities in Manila, with at least two months of clinical exposure and prior knowledge or experience with GenAI, participated in four online focus group discussions via Google Meet. Data were analyzed using thematic analysis. Four major themes emerged: (1) GenAI functions as a helpful support tool, (2) perceived limitations of GenAI integration, (3) potential risks of GenAI, and (4) the status of GenAI integration. Interns reported using GenAI to support evaluation and intervention planning, documentation, and academic learning, contributing to perceived improvements in confidence, efficiency, and clinical preparation. However, they emphasized that GenAI cannot replace OT clinical reasoning because of occasional inaccuracies, limited contextualization, and difficulty capturing holistic, client-centered OT practice. Institutional and contextual barriers, particularly variable guidelines and limited relevance to the Filipino practice context, further shaped its use. Ethical concerns included risks to patient data, confidentiality, safety, and over-reliance, which may affect clinical reasoning and professional development. Despite these concerns, interns viewed GenAI integration as unavoidable and recommended structured training, clearer policies, supervised practice opportunities, and ethical guidance. GenAI is perceived as a valuable yet conditional support tool in OT clinical training. Responsible integration requires alignment with professional standards, institutional safeguards, and continued emphasis on clinical reasoning, ethical practice, and contextually responsive OT education.

**Keywords:** generative artificial intelligence; occupational therapy; clinical training; interns; clinical education; Philippines

## References

[1] Kaelin, V. C., Valizadeh, M., Salgado, Z., Sim, J. G., Anaby, D., Boyd, A. D., et al. (2022). Capturing and operationalizing participation in pediatric re/habilitation research using artificial intelligence: A scoping review. *Frontiers in Rehabilitation Sciences, 3*. https://doi.org/10.3389/fresc.2022.855240

[2] Kaelin, V. C., Nilsson, I., & Lindgren, H. (2024). Occupational therapy in the space of artificial intelligence: Ethical considerations and human-centered efforts. *Scandinavian Journal of Occupational Therapy, 31*(1), Article 2421355. https://doi.org/10.1080/11038128.2024.2421355

[3] Lee, S. A., & Park, J. H. (2025). Artificial intelligence in occupational therapy documentation: Chatbot vs. occupational therapists. *DIGITAL HEALTH, 11*. https://doi.org/10.1177/20552076251386657

# Subacromial Region Ultrasound Image Enhancement and Segmentation for Acromiohumeral Distance Measurement

**Engr. Glenn C. Virrey, MSc, ECT, Gwen Stefani C. Loreno, Sean Darrel Y. Mariano, John Angelo B. Tabang, Jerald Tyrone V. Tadena, Lance Adriane B. Wallit, Prof. Consuelo G. Suarez, MD, PhD, FPARM, Jan-Tyrone C. Cabrera, MD, FPARM, Mary Monica B. Bueno, MD, PHD, Asst. Prof. Maria Belinda Cristina C. Fidel, MD, FPARM**

*Graduate School, University of Santo Tomas, España, Manila, Philippines*
*Faculty of Engineering, University of Santo Tomas, España, Manila, Philippines*
*College of Rehabilitation Sciences, University of Santo Tomas, España, Manila, Philippines*
**Corresponding author: gcvirrey@ust.edu.ph*

Shoulder Impingement Syndrome (SIS) is a prevalent musculoskeletal disorder characterized by the compression of soft tissues within the subacromial space. The Acromiohumeral Distance (AHD) serves as a critical diagnostic parameter for SIS. This study proposes an automated image processing framework to enhance ultrasound (US) clarity, segment the subacromial region, and measure AHD. A multi-stage enhancement pipeline was developed, with the "Enhance6" configuration utilizing adaptive histogram equalization, low boundary leveling, and 20 iterations of gamma correction emerging as the optimal protocol. Results demonstrate a 90% improvement in Signal-to-Noise Ratio (SNR), increasing from 3.3 to 6.6. Notably, the framework intentionally reduced the Contrast-to-Noise Ratio (CNR) to preserve the natural appearance of bone interfaces, a feature clinically validated to improve anatomical landmark identification. The automated measurement system achieved strong agreement with manual methods, reducing mean absolute error by approximately 60%, providing a practical, low-cost diagnostic support tool for resource-constrained environments.

# A Comprehensive Preprocessing Framework for Automated Handwriting Evaluation Aligned with the Minnesota Handwriting Assessment

**J.T. Amores,[1,2] A.C. Frias,[1,2] P.M. Espino,[3] I.B. Gomez,[3] and E.A. Roxas[2]**

*[1]Graduate School, University of Santo Tomas, España, Manila, Philippines*

*[2]Faculty of Engineering, University of Santo Tomas, España, Manila, Philippines*

*[3]College of Rehabilitation Sciences, University of Santo Tomas, España, Manila, Philippines*

**Corresponding author: josephden.amores.gs@ust.edu.ph*

Accurate handwriting assessment in children is essential for evaluating fine motor skills, cognitive development, and academic readiness. Standardized tools such as the Minnesota Handwriting Assessment (MHA) provide structured criteria for evaluating legibility, form, alignment, size, and spacing, yet manual scoring is subjective, time-consuming, and challenging to scale across large and marginalized populations. This study presents a comprehensive preprocessing framework that serves as a foundation and offers support for automated handwriting evaluation directly aligned with MHA standards. Handwritten samples from children aged 6–10 years were captured using a reproducible prototype comprising an LED-illuminated studio softbox and camera, ensuring consistent orientation, lighting, and high-quality image acquisition across diverse participants. The preprocessing pipeline includes region-of-interest cropping, grayscale conversion, non-local means denoising, adaptive thresholding, guideline removal, contour-based noise filtering, and morphological refinement, producing clean, high-contrast images of individual letters. Interactive letter cropping ensures precise isolation of each character, preserving structural integrity critical for evaluating MHA-specific handwriting components. Preliminary results demonstrate that the pipeline effectively reduces background noise, compensates for illumination and ink variations, and maintains fine motor details necessary for scoring MHA criteria. By producing standardized letter-level inputs, this framework provides a foundation for scalable and objective automated evaluation, enabling future systems to deliver interpretable, component-wise assessments consistent with MHA scoring guidelines. This methodology provides a novel, fundamental, practical approach for enhancing accessibility, objectivity, and efficiency in educational and clinical handwriting assessment.

# EXPLORING THE INTEGRATION OF ARTIFICIAL INTELLIGENCE TOOLS TO ENHANCE COUNSELING PRACTICES AMONG FILIPINO OCCUPATIONAL THERAPISTS

J.G. Cua, M.S. Tan[1], G.B. Evangelista[2], M.L. Chavez[2]

*Department of Occupational Therapy, University of Santo Tomas, Manila, Philippines [1]*

*Department of Psychology, Miriam College, Manila, Philippines [2]*

*Department of Education, Miriam College, Manila, Philippines [3]*

**Corresponding author: jcgurtiza@ust.edu.ph*

Artificial intelligence (AI) is increasingly utilized in healthcare; however, its integration into occupational therapy (OT) counseling remains insufficiently examined within the Philippine context. This study investigated the ethical and contextual integration of AI into OT counseling practice and education through a Participatory Action Research approach. Data collection involved a focus group discussion with licensed occupational therapists and five key informant interviews with educators, clinicians, and a school-based guidance counselor. Participant-validated transcripts were analyzed using Reflexive Thematic Analysis, following Braun and Clarke's six-phase framework. The analysis identified five themes: AI awareness, barriers and enablers to integration, applications in counseling-oriented OT practice, and recommendations for ethical use. Results suggest that Filipino occupational therapists possess emerging yet limited familiarity with AI, coupled with a strong interest in structured guidance and safeguards. The study developed an AI Protocol Manual to facilitate ethical and culturally responsive integration of AI in OT counseling, underscoring that AI should augment rather than replace the human-centered foundations of practice.

**Keywords:** artificial intelligence, occupational therapy, counseling, ethics, education, participatory action research, Philippines

## Methodology

***Figure 1***

Participatory Action Research Framework for the Development of an AI Protocol for Filipino Occupational Therapy




## Results

***Figure 2***

**Conceptual Map**

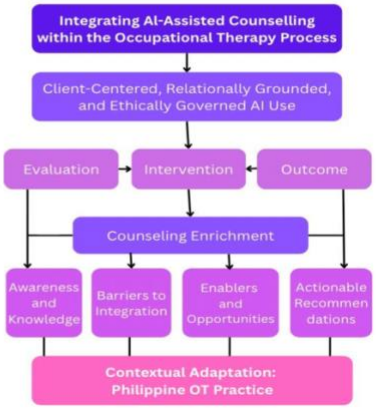



*Figure 2* serves as a visual synthesis of the study's findings, illustrating that the Integration of AI into occupational therapy must occur within a cyclical and relational process anchored in counseling. Rather than a linear or mechanistic tool, AI operates within and across the stages of the OT process. During evaluation, AI supports information gathering, reflection, and data synthesis. Within the intervention, it assists in generating counseling resources, tracking client progress, and structuring reflective dialogue. In outcomes, it enables monitoring, documentation, and adaptive planning. Surrounding this core cycle are the five interrelated themes, each representing a developmental condition that enables ethical and contextual Integration. Together, these components form a reflexive ecosystem of augmented practice—a model where echnology, therapist judgment, and client participation co-evolve to strengthen the therapeutic alliance and promote culturally grounded, client-centered care.

Implications

**Implications for Counseling Practice in Occupational Therapy**

This study reaffirms that counseling is not a supplementary activity in occupational therapy (OT) but a central, relational process that sustains empathy, ethical reflection, and client-centered care, and integrating artificial intelligence (AI) into counseling-oriented OT practice requires a reframing of technology's purpose—from a mechanistic substitute for reasoning to a reflective partner that enhances therapeutic presence and attunement. Participants consistently described AI as a cognitive aid that helps organize thoughts, summarize reports, and manage documentation, allowing therapists to devote more time and attention to the relational aspects of client care. However, the value of AI integration depends on maintaining counseling ethics as its foundation. Practitioners emphasized that empathy, presence, and relational integrity cannot be automated; they require genuine human interaction. Counseling principles—trust, authenticity, and mutual respect—must continue to guide the therapist's reflective reasoning. In this sense, AI functions as a cognitive amplifier, enhancing the therapist's ability to reflect, plan, and evaluate, but never replacing the human process of therapeutic understanding. Also, the use of AI must therefore remain transparent, ethically supervised, and guided by professional discernment to ensure that the therapist's presence remains at the core of the therapeutic alliance.

**Implications for Model Building**

The study contributes to model building by proposing an integrative conceptual framework that positions counseling at the center of the OT process while embedding AI as a supportive and adaptive layer. The model illustrated in Figure 2: *Conceptual Integration Model*, demonstrates how counseling bridges the phases of the Occupational Therapy Practice Framework (AOTA, 2020)—evaluation, intervention, and outcomes—while AI serves as an assistive layer that supports reflection, ethical reasoning, and clinical decision-making. The model reframes counseling as a cross-cutting mechanism in the OT process rather than a distinct intervention category. It functions as the interpretive and relational thread that connects data with human meaning. The findings suggest that AI can operationalize what Topol (2019) refers to as *augmented intelligence*—a form of collaboration between human insight and machine efficiency that enhances, rather than replaces, professional judgment. Within this paradigm, AI's role is reflective and supportive, not directive or deterministic. The Filipino adaptation of the model contextualizes global AI ethics within local values of *pakikipagkapwa* (shared humanity) and *bayanihan* (collective responsibility). These cultural foundations affirm that technological adoption must preserve the relational and communal orientation of care that defines Filipino counseling and OT practice. Model development,

therefore, should continue to integrate cultural sensitivity and ethical literacy as operational dimensions of AI-assisted occupational therapy. This integrative framework can guide the development of future protocols, curriculum reforms, and practice guidelines that reflect both global standards and local contexts, thereby enhancing the effectiveness of healthcare delivery.

**Implications for Occupational Therapy Education**

There is a pressing need to modernize occupational therapy education in response to the digital transformation in health and rehabilitation. The current CHED Memorandum Order (CMO) No. 52, Series of 2017, defines program outcomes related to ethical reasoning and reflective practice but omits competencies related to AI, digital health, and data literacy. Participants' reflections show that most practitioners and educators learned about AI independently, outside formal instruction, often through self-exploration and informal experimentation. Integrating AI literacy into OT curricula must go beyond technical instruction. It should emphasize the reflective and ethical use of technology in counseling contexts. Embedding digital ethics, simulation-based learning, and case-based reflection into psychosocial and counseling-related courses can enhance students' readiness to navigate complex, technology-assisted therapeutic scenarios. Interdisciplinary collaboration among OT, psychology, computer science, and bioethics faculties can further strengthen the educational infrastructure for the responsible use of AI. Faculty development and continuing professional education are equally vital. Continuing Professional Development (CPD) programs should include training in AI ethics, data governance, and applied counseling supervision to ensure practitioners remain adaptable and ethically grounded throughout their careers. This approach aligns with Knowles' (1984) principles of adult learning, which emphasize self-directed, experiential, and contextually relevant learning as essential to professional growth.

**Implications for Policy and Institutional Governance**

At the policy level, the findings underscore the need for institutional frameworks that ensure the integration of AI remains ethical, transparent, and contextually appropriate. Professional bodies such as the Philippine Academy of Occupational Therapy (PAOT) and the Professional Regulation Commission (PRC) are in strategic positions to establish national guidelines that regulate the ethical use of AI in OT practice. Participants repeatedly called for formal policies, emphasizing that "We need clear guidelines on how to use AI ethically" (KII Participant, 2025).

Policy initiatives should address data privacy, algorithmic accountability, and informed consent, in line with the Philippine Data Privacy Act of 2012 (RA 10173) and the World Health Organization's (2021) principles for AI governance. Health and academic institutions can establish AI Ethics Review Boards to monitor implementation, assess risk, and guide ethical decision-making in practice and education. Additionally, AI-related competencies can be embedded in accreditation, licensure, and CPD frameworks, ensuring that ethical and technological literacy become hallmarks of professional competence. Institutional partnerships with technology developers and universities should be encouraged to co-create culturally responsive AI tools that align with Filipino psychosocial contexts. This collaborative, multi-sectoral approach can help strike a balance between innovation and ethical stewardship, ensuring that technological advancements serve the greater good of human well-being.

**Implications for Further Research**

The study identifies several directions for future research on integrating artificial intelligence (AI) into counseling-oriented occupational therapy (OT). Building on current qualitative findings and the development of the AI protocol manual, subsequent research should progress to Phase 6: Implementation and Evaluation, which remains unaddressed in this investigation. Future studies should examine the real-world application of the protocol manual in both academic and clinical OT settings through rigorously designed pilot implementations. These studies may employ quantitative and mixed-methods approaches to assess the feasibility, adaptability, and outcomes of AI-supported counseling practices.

When evaluating implementation, future research should consider metrics such as practitioner adherence to the protocol's ethical guidelines, including consistent use of AI-specific informed consent procedures and completion of ethical checklists before AI-supported sessions. Process indicators may include the frequency and type of AI-supported activities adopted, such as documentation support, psychoeducational material generation, and counseling preparation. Additional measures may include changes in time allocation between administrative tasks and direct client interaction, as well as practitioner self-reported confidence and ethical comfort in using AI-assisted tools. These indicators can clarify whether AI integration meaningfully supports counseling practice without displacing professional judgment or relational engagement.

At the client level, future research should incorporate evaluation measures related to counseling quality and relational outcomes. These measures may include client-reported understanding of AI use, perceived transparency and trust, comfort with AI-supported materials, and perceived impact on engagement and participation in therapy. Importantly, consistent with the study's Participatory Action Research (PAR) orientation, clients may be actively involved in co-defining what transparency and trust mean in AI-assisted counseling contexts, rather than treating these constructs solely as predefined variables. Through collaborative feedback sessions, reflective interviews, or participatory evaluation activities, clients can articulate the indicators of ethical clarity, relational safety, and openness that are most meaningful to them. Where appropriate, brief assessments of therapeutic alliance stability or qualitative feedback on relational safety may be used to examine whether AI integration maintains the counseling values central to occupational therapy practice. Collectively, these collaboratively informed metrics would enable empirical testing of the protocol manual's underlying assumptions and demonstrate its practical and relational value across diverse OT environments.

Further research should address Phase 7: Dissemination and Policy Recommendations, which was intentionally beyond the scope of the current study. This phase may involve distributing the finalized AI protocol manual to academic institutions, professional organizations, and policymakers, followed by systematic examination of its influence on curriculum development, faculty training, continuing professional development programs, and ethical governance related to AI use in OT counseling. Policy-oriented studies should assess institutional uptake of protocol recommendations, alignment with national educational and professional standards, and the establishment of governance mechanisms for AI vetting, data privacy, and ethical oversight.

Incorporating client and family perspectives is essential for understanding how AI-mediated counseling practices influence perceptions of empathy, trust, and relational safety. Additional research focusing on educator readiness, institutional capacity, and resource availability can further inform curriculum reforms that integrate AI ethics, reflective counseling practice, and culturally responsive competencies. Consistent with the study's Participatory Action Research (PAR) foundation, future investigations should co-design, implement, and evaluate AI-supported counseling practices with Filipino occupational therapists, educators, and clients to ensure cultural, linguistic, and ethical relevance.

References

Abd-Alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, *132*, 103978. https://doi.org/10.1016/j.ijmedinf.2019.103978

Aguas, J. J. S. (2016). The Filipino Value of Pakikipagkapwa-Tao Vis-À-Vis Gabriel Marcel's Notion of Creative Fidelity and Disponibilitè. *Scientia: Research Journal of the College of Arts and Sciences, San Beda University*.

Ahuja, A. S. (2019). The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ*, *7*, e7702. https://doi.org/10.7717/peerj.7702

Aljerian, N., Arafat, M., Aldhubib, A., Almohaimeed, I., Alsultan, A., Alhosaini, A., Aloqayli, A., AlRabiah, B., AlBanyan, S., Ashry, L. B., AlGhassab, R., Alzahrani, R., Alhenaki, G., Alnazzawi, E., AlMughaiseeb, A., & Alanazi, A. (2022). Artificial

Intelligence in Health care and its application in Saudi Arabia. *International Journal of Innovative Research in Medical Science*, *7*(11), 666–670. https://doi.org/10.23958/ijirms/vol07-i11/1558

American Occupational Therapy Association. (2020). Occupational therapy practice framework: Domain and process (4th ed.). *American Journal of Occupational Therapy*, *74*(Suppl. 2), 7412410010. https://doi.org/10.5014/ajot.2020.74S2001

Aponte, H. J. (2021). The soul of Therapy: the therapist's use of self in the therapeutic relationship. *Contemporary Family Therapy*, *44*(2), 136–143. https://doi.org/10.1007/s10591-021-09614-5

Arushi, Dillon, R., Teoh, A. N., & Dillon, D. (2022). Voice Analysis for stress detection and application in virtual Reality to improve public speaking in real-time: A review. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2208.01041

Avello-Sáez, D., & Estrada-Palavecino, L. (2023). ChatGPT and its impact on competence training in occupational therapists: a reflection on academic integrity. *Cadernos Brasileiros de Terapia Ocupacional*, *31*. https://doi.org/10.1590/2526-8910.ctoen271035342

B, N. P. K., B, N. S., & Patrer, N. D. (2020). Emotion recognition with AI: Techniques and applications. *World Journal of Advanced Research and Reviews*, *8*(2), 344–352. https://doi.org/10.30574/wjarr.2020.8.2.0362

Bankins, S., Hu, X., & Yuan, Y. (2024). Artificial intelligence, workers, and future of work skills. *Current Opinion in Psychology*, *58*, 101828. https://doi.org/10.1016/j.copsyc.2024.101828

Basu, K., Sinha, R., Ong, A., & Basu, T. (2020). Artificial intelligence: How is it changing medical sciences and its future? *Indian Journal of Dermatology*, *65*(5), 365. https://doi.org/10.4103/ijd.ijd_421_20

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage High-Risk and High-Cost patients. *Health Affairs*, *33*(7), 1123–1131. https://doi.org/10.1377/hlthaff.2014.0041

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

https://doi.org/10.1191/1478088706qp063oa Chandler, D., & Torbert, B. (2003). Transforming inquiry and action. *Action Research*, *1*(2), 133–152. https://doi.org/10.1177/14767503030012002

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, *4*(2), e19. https://doi.org/10.2196/mental.7785

Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). SAGE Publications.

Garcia, J. A., Lorenzo, F. M., & David, E. J. R. (2021). Occupational therapy in the Philippines: Challenges and opportunities in post-disaster rehabilitation. *Philippine Journal of Occupational Therapy*, *15*(1), 12-25.

Gutierrez, K. L. T., & Viacrusis, P. M. L. (2023). Bridging the gap or widening the divide: A call for Capacity-Building in artificial intelligence for healthcare in the Philippines. *Journal of Medicine University of Santo Tomas*, *7*(2), 1325–1334. https://doi.org/10.35460/2546-1621.2023-0081

Habermann, J. (2021). *Psychological Impacts of COVID-19 and Preventive Strategies: A review*. Humanities Publishing House. https://doi.org/10.13140/RG.2.2.29662.15681

Haglund, L., Thorell, L. H., & Walinder, J. (2000). Occupational therapy and the therapeutic use of self. *Scandinavian Journal of Occupational Therapy*, *7*(1), 11-16. https://doi.org/10.1080/110381200443572

Haitham Jahrami. (2023). Unlocking the potential: Exploring the opportunities for occupational therapists in cyberpsychology. *Hong Kong Journal of Occupational Therapy*. https://doi.org/10.1177/15691861231222074

Hisler, G. C., Young, K. S., Cumpanasoiu, D. C., Palacios, J. E., Duffy, D., Enrique, A., Keegan, D., & Richards, D. (2024). Incorporating a deep-learning client outcome

prediction tool as feedback in supported internet-delivered cognitive

behavioural therapy for depression and anxiety: A randomised controlled trial within routine clinical practice. *Counselling and Psychotherapy Research*. https://doi.org/10.1002/capr.12771

Kinsella, E. A. (2020). Professional identity in occupational therapy: A constructivist grounded theory study. Scandinavian Journal of Occupational Therapy, 27(4),

251-262. https://doi.org/10.1080/11038128.2019.1577168
Iannone, A., & Giansanti, D. (2023). Breaking Barriers—The Intersection of AI and

Assistive Technology in Autism Care: A Narrative review. *Journal of Personalized Medicine*, *14*(1), 41. https://doi.org/10.3390/jpm14010041 Johnson, D., Deterding, S., Kuhn, K. A., Staneva, A., Stoyanov, S., & Hides, L. (2016).

Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions*, *6*, 89-106. https://doi.org/10.1016/j.invent.2016.10.002

Kinsella, E. A. (2020). Professional identity in occupational therapy: A constructivist grounded theory study. *Scandinavian Journal of Occupational Therapy*, *27*(4), 251-262. https://doi.org/10.1080/11038128.2019.1577168

Lima, R., Chirico, A., Varandas, R., Gamboa, H., Gaggioli, A., & Badia, S. B. I. (2024). Multimodal emotion classification using machine learning in immersive and non-immersive virtual reality. *Virtual Reality*, *28*(2). https://doi.org/10.1007/s10055-024-00989-y

Lorenzo, F. M., Garcia, J. A., & Enriquez, V. G. (2018). Bayanihan in occupational therapy: Community-based rehabilitation in the Philippines. *Asian Journal of Occupational Therapy*, *14*(2), 45-52.

Luxton, D. D. (2020). Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization*, *98*(4), 285-287. https://doi.org/10.2471/BLT.19.237636

Miner, A. S., Milstein, A., & Hancock, J. T. (2017). Talking to machines about personal mental health problems. *JAMA*, *318*(13), 1217. https://doi.org/10.1001/jama.2017.14151

Naderbagi, A., Loblay, V., Zahed, I. U. M., Ekambareshwar, M., Poulsen, A., Song, Y. J. C., Ospina-Pinillos, L., Krausz, M., Kamel, M. M., Hickie, I. B., & LaMonica, H. M. (2024). Cultural and Contextual adaptation of Digital Health Interventions: Narrative review. *Journal of Medical Internet Research*, *26*, e55130. https://doi.org/10.2196/55130

Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., & Eberhardt, J. (2024). Enhancing mental health with Artificial Intelligence: Current trends and future prospects. *Journal of Medicine Surgery and Public Health*, *3*, 100099. https://doi.org/10.1016/j.glmedi.2024.100099

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., ... & UnÜtzer, J. (2020). The Lancet Commission on global mental health and sustainable development. The Lancet, 392(10157), 1553-1598. https://doi.org/10.1016/S0140-6736(18)31612-X

Parsons, T. D., & Rizzo, A. A. (2007). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(3), 250–261. https://doi.org/10.1016/j.jbtep.2007.07.007

Petersson, L., Larsson, I., Nygren, J., Nilsen, P., Neher, M., Reed, J., Tyskbo, D., & Svedberg, P. (2022). Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. *BMC Health Services Research*, *22*(1). https://doi.org/10.1186/s12913-022- 08215-8

Reason, P., & Bradbury, H. (2008). *The Sage handbook of action research : participative inquiry and practice*. In Sage eBooks. https://jprm.scholasticahq.com/article/13244-participatory-research-methods- choice-point s-in-the-research-process

Reason, P., & Torbert, W. (2001). The action turn. *Concepts and Transformation*, *6*(1), 1– 37. https://doi.org/10.1075/cat.6.1.02rea

Reddy, Sandeep. (2021). Artificial intelligence and healthcare—why they need each other?. *Journal of Hospital Management and Health Policy*, *5*. 10.21037/jhmhp-2020-ai-03.

Reyes, J. (2015). LoóbandKapwa: An Introduction to a Filipino Virtue Ethics. *Asian Philosophy*, *25*(2), 148–171.

https://doi.org/10.1080/09552367.2015.1043173

Rogers, C. R. (1951). *Client-centered Therapy*. Boston: Houghton Mifflin.
Ross, M. K., Wei, W., & Ohno-Machado, L. (2014). “Big data” and the electronic health

record. *Yearbook of Medical Informatics*, *23*(01), 97–104.

https://doi.org/10.15265/iy-2014-0003

Schreier, M. (2012). Qualitative content analysis in practice. SAGE Publications. Soriano, C. R., David, C. C., & Torres, M. T. (2021). Trust in technology: A study of

Filipino attitudes toward AI in healthcare. *Philippine Journal of*

*Psychology*, *54*(1), 78-95.
Tanaka, G., Inadomi, H., Kikuchi, Y., & Ohta, Y. (2020). Evaluating an artificial

intelligence-powered mental health app in Japan: A feasibility study. Asian Journal of Psychiatry, 52, 102025. https://doi.org/10.1016/j.ajp.2020.102025

Taylor, R. R., Lee, S. W., Kielhofner, G., & Ketkar, M. (2009). Therapeutic use of self: A nationwide survey of practitioners’ attitudes and experiences. *American Journal of Occupational Therapy*, *63*(2), 198-207. https://doi.org/10.5014/ajot.63.2.198

114

Topol, E. J. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.

Vaughn, L. M., & Jacquez, F. (2020). Participatory Research Methods – Choice Points in the Research Process. *Journal of Participatory Research Methods*, *1*(1). https://doi.org/10.35844/001c.13244

Witty, M. C. (2006). Client-Centered therapy. In *Springer eBooks* (pp. 35–50). https://doi.org/10.1007/978-0-387-29681-4_3

World Health Organization. (2020). *Global strategy on human resources for health: Workforce 2030*. WHO Press.

**PRESENTATION TYPE**

☐ Oral Presentation
✔ Poster Presentation

**CONFERENCE TOPIC AREA**

☐ Social Inequality & Economic Disparities
☐ Public Health & Well-being
✔ Education & Social Change
☐ Environmental & Sustainability Issues
☐ Technology & Society
☐ Conflict Resolution & Human Rights
☐ Governance & Policy Making
☐ Artificial Intelligence
☐ Wireless Technologies
☐ Cyber Security
☐ Sustainable Technology & Emerging Technologies
☐ Other (please specify): _____

# Practices Of Filipino Speech-Language Pathologists Based In Metro Manila In Managing Pragmatic Difficulties Among Children With Autism Spectrum Disorder

**Bautista, Elinor C.; Esguerra, Mary Ann Gisselle O.; Mabanag, Maria Kiana; Adriatico, Anekiel Jayne; Dulay, Bianca Laine; Medallo, Angela Kyle; Tumang, Alexandra Nicole**

*Department of Speech-Language Pathology, University of Santo Tomas, Manila, Philippines*
*College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*
** Corresponding author: elcbautista@ust.edu.ph*

Pragmatic language skills are essential for meaningful social communication yet remain particularly challenging for children with Autism Spectrum Disorder (ASD).Despite the growing prevalence of ASD in the Philippines, no published research exists on the standards of practice for pragmatic assessment and intervention among Filipino speech-language pathologists (SLPs). This study aimed to describe the practices of Filipino SLPs in private clinics within the National Capital Region (NCR) in managing pragmatic language difficulties of school-aged children with ASD. A qualitative descriptive design was employed. Eight registered Filipino SLPs with at least five years of clinical experience participated in one-on-one semi-structured interviews conducted online. Data were analysed through Braun and Clarke's Six-Phase Thematic Analysis using NVivo software. Four major themes emerged: assessment, intervention planning, implementation, and collaborative efforts. Participants reported utilising formal assessment tools primarily to evaluate prerequisite receptive, expressive, and higher language skills, while informal and criterion-referenced approaches were preferred for assessing pragmatics due to the absence of culturally appropriate norms for Filipino children. Intervention planning was highly individualized and guided by the child's language abilities, interests, and social environment. Commonly utilised approaches included Social Thinking and peer-mediated intervention, although the need for cultural adaptation of Western-based frameworks was noted. Therapy sessions typically began through individualized sessions before transitioning to dyad or group settings using real-life, context-driven activities. Intervention activities included roleplay, social stories, videos, board games, and context-based conversational tasks designed to reflect real-life situations. Interdisciplinary collaboration with teachers, occupational therapists, and parents was considered essential for goal-setting, behavioural support, and carryover across settings. The findings suggest that Filipino SLPs' practices in managing pragmatic difficulties are context-driven, culturally responsive, and grounded in clinical reasoning, underscoring the need for locally developed assessment frameworks and intervention guidelines for children with ASD in the Philippine context.

**Keywords:** Practices, Autism Spectrum Disorder (ASD), Filipino Speech-Language Pathologists (SLPs), managing pragmatic difficulties, private clinics, Metro Manila

# Children's Screen Time and Its Correlation with Well-being, Parental Attitudes, and Guardians' Screen Time in Filipino Children in Malabon City: An Observational Cross-Sectional Study

**Espino, Paulin Grace M.[1,2]; Tuazon, John Alexander David B.[2]; Mortos, Mikaela Kassy O.[2]; Anteojo, Nashley Gayle DC.[2]; Nobleza, Kervin Joshua O.[2]; Pangan, Cestel Margaret R.[2]; Precioso, Maria Pauline G.[2]; Vargas, Marianne C.[2]**

[1] Center for Health Research and Movement Sciences, University of Santo Tomas, Manila, Philippines

[2] Department of Occupational Therapy, College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines

***** *Corresponding author: pmespino@ust.edu.ph*

The global trend toward increased screen use has led to a surge in screen time among young children, prompting attention to its relationship with children's well-being. Well-being includes physical, emotional, psychological, and social health. In the Philippines, where access to technology is growing rapidly, it is important to examine how it affects young children. This study aimed to: (1) determine the correlation between screen time and the well-being of young children, (2) identify the correlation between parental attitudes and children's screen time, and (3) identify the correlation between children's and their guardians' screen time. A cross-sectional, observational, and analytical research design was used in Filipino children in Kindergarten to Grade 2 enrolled in schools in Malabon City, Metro Manila. The Media Exposure Survey and the Filipino Child Well-Being were used to measure the study's variables. The Mann–Whitney U test was used to compare public and private school variables. A Spearman's rank-order correlation was performed to assess the relationship between the variables. No significant correlation was observed between children's screen time and well-being, or parental attitudes and children's screen time. A strong positive correlation was found between children's and guardians' time spent on screens ($r_s = 0.67$, $p < .001$). A very strong correlation between children's and guardians' screen time was found for Kindergarten students ($r_s = 0.80$, $p < .001$). The results indicate that the amount of time spent on screen use is not sufficient to determine correlation with child well-being and parental attitudes, implying the importance of screen use moderation rather than avoidance. The strong correlation between children's and their guardians' screen time suggests the importance of modeling in children's screen use. Understanding factors that contribute to guardians' excessive screen use remains important for future research, particularly in identifying supports that can help families develop better screen habits.

**Keywords:** Screen time, Child well-being, Parental attitudes, Screen type, Filipino children

# iSULAT: Feature Definition and Validation of a Sensor-Based Pen for Handwriting Analysis

A.C. Frias,[1,2] J.T. Amores,[1,2] P.M. Espino,[3] I.B. Gomez,[3] and E.A. Roxas[2]

[1]*Graduate School, University of Santo Tomas, España, Manila, Philippines*

[2]*Faculty of Engineering, University of Santo Tomas, España, Manila, Philippines*

[3]*College of Rehabilitation Sciences, University of Santo Tomas, España, Manila, Philippines*

**Corresponding author: antoniomiguel.frias.gs@ust.edu.ph*

Handwriting is a fundamental fine motor skill in early childhood education, with school-aged children spending a substantial portion of their academic activities engaged in writing-related tasks. Despite its importance, handwriting difficulties are reported in a considerable proportion of school-aged children, often affecting academic performance, confidence, and learning development. Existing handwriting assessment approaches rely heavily on expert observation and structured evaluation procedures, which can be time-intensive and limit opportunities for continuous and objective measurement. These challenges highlight the need for supportive tools that enhance and complement clinical assessment practices. This study presents the feature definition and validation of a sensor-based pen system, iSULAT (Intelligent Stroke Utilization, Learning, Assessment, and Testing), where "isulat" is the Filipino term for "to write", reflecting the system's focus on handwriting-centered analysis. The system integrates motion, pressure, and orientation sensors to acquire raw handwriting signals, from which thirteen handwriting features are formally defined, encompassing spatial, temporal, kinematic, pressure-related, and orientation-based characteristics of handwriting execution. Feature validity is established through multiple complementary approaches, including construct-based analysis, reliability assessment across repeated trials, expert validation by clinicians in the field of rehabilitation sciences, and verification using an independent camera-based recording setup to confirm temporal and spatial measurement accuracy. Results demonstrate stable feature behavior, expected inter-feature relationships, and consistency between sensor-derived measurements and observable writing events. By establishing a validated and interpretable set of handwriting features, this work provides a strong foundation for objective handwriting analysis and supports the development of technology-assisted tools that augment handwriting assessment and monitoring.

**PRESENTATION TYPE**

- [x] Oral Presentation
- [ ] Poster Presentation

**CONFERENCE TOPIC AREA**

- [ ] Social Inequality & Economic Disparities
- [ ] Public Health & Well-being
- [ ] Education & Social Change
- [ ] Environmental & Sustainability Issues
- [ ] Technology & Society
- [ ] Conflict Resolution & Human Rights
- [ ] Governance & Policy Making
- [ ] Artificial Intelligence
- [ ] Wireless Technologies
- [ ] Cyber Security
- [x] Sustainable Technology & Emerging Technologies
- [ ] Other (please specify): _____

# Development of a Thermo-Tactile Stimulation Device with EEG-based Neurofeedback for Stress Intervention

**Ocenar, Samuel John A.**[1,2,*]; **Fabay, Lianna Mae A.**[1,2]; **Danganan, Lester Mico L.**[5]; **Nieles, Jay Patrick M.**[1]; **Catindig-Dela Rosa, Joseree Ann S.**[6,7]; **Prado, Seigfred V.**[1,2,3,4]

[1] Brain Research Advancement and Innovations in Neurotechnology (BRAIN) Lab, University of Santo Tomas, Manila, Philippines

[2] Graduate School, University of Santo Tomas, Manila, Philippines

[3] Department of Electronics Engineering, University of Santo Tomas, Manila, Philippines

[4] Research Center for Engineering and Technological Sciences, University of Santo Tomas, Manila, Philippines

[5] NSW Health, Sydney, New South Wales, Australia

[6] Faculty of Medicine and Surgery, University of Santo Tomas, Manila, Philippines

[7] National Center for Geriatric Health, Manila, Philippines

* *Corresponding author: samueljohn.ocenar.gs@ust.edu.ph*

Stress is a pervasive concern among young adults aged 18 to 25, with university students particularly vulnerable due to academic, social, and financial pressures. Chronic stress impairs emotional regulation and cognitive function, increasing the risk for anxiety and depression. Although pharmacological and cognitive-behavioral interventions are available, their use may be limited by side effects, accessibility barriers, and high cognitive demands. Sensory-based interventions offer a low cognitive demand alternative by engaging peripheral pathways involved in relaxation. Thermal stimulation provides a gradual soothing effect by activating thermosensory pathways associated with comfort, safety, and autonomic regulation, while tactile vibration provides a more immediate sensory input through mechanoreceptor pathways linked to parasympathetic modulation and emotional regulation. However, when used independently alone, thermal stimulation may have a slower perceptual onset, while tactile stimulation may lack sustained affective depth. Combining both modalities may therefore provide a broader, more adaptable sensory experience by integrating the emotional warmth of thermal input with the dynamic regulatory effects of tactile stimulation. In this study, a non-invasive multimodal scalp stimulation device was developed to deliver localized thermal and tactile stimulation inputs while allowing simultaneous electroencephalogram (EEG) recording. The device used Peltier modules for thermal stimulation at 39°C, 40°C, and 41°C, coin-type or linear resonant actuators for tactile stimulation at 20, 30, and 40 Hz, and an ESP32-based control architecture with supporting components, including MOSFETs, resistors, drivers, and temperature-monitoring sensors. Participants undergo baseline EEG recording, cognitive stress induction using

Stroop and arithmetic tasks, post-stress EEG recording, eight-minute stimulation exposure, and post-intervention EEG recording with subjective stress and comfort ratings. EEG responses were evaluated through spectral and nonlinear analyses to determine neurophysiological changes associated with thermal-only, tactile-only, placebo, and combined stimulation conditions. This study positions EEG-guided thermo-tactile stimulation as a personalized, non-invasive, and scalable approach for stress intervention in university populations.

**Keywords:** Multisensory Stimulation, EEG-based Neurofeedback, Thermo-Tactile Integration, Stress Modulation, Personalized Intervention

## References

[1] Brelet, L., & Gaffary, Y. (2022). EEG-based assessment of stress modulation using sensory stimuli. *International Journal of Psychophysiology, 177*, 30 to 37.
[2] Yang, C., et al. (2023). Thermal stimulation in stress modulation: An EEG study. *Biological Psychology, 176*, 108587.
[3] Lopes Alves, L., et al. (2025). Electrical stimulation and vibration therapy for neuromodulation. *Biomedical Engineering Online, 24*(1), Article 12.
[4] Zhou, L., et al. (2023). Combining vibration and warmth for emotion regulation: A pilot study. *IEEE Transactions on Affective Computing*.
[5] Stefaniak, A., et al. (2021). Mapping stress-responsive brain regions: A neuroacupuncture approach. *Brain Sciences, 11*(12), 1612.

# Factors Associated with Dietary Supplement Use Among Filipino National Athletes: A Cross-Sectional Study

**Asst. Prof. Karen Leslie Pineda[1*], MSPH, Catherine Joy Escuadra, PhD[2]**

[1]Department of Sports Science, University of Santo Tomas, Manila, Philippines
[2]Department of Physical Therapy, University of Santo Tomas, Philippines
[1] * *Corresponding author: klpineda@ust.edu.ph*

Dietary supplement use among athletes has become increasingly prevalent across various sports disciplines; however, inconsistencies in definitions and measurement approaches have led to conflicting evidence. Although international studies have examined supplement use, limited evidence exists on its prevalence among Filipino national athletes. Understanding the prevalence and factors associated with dietary supplement use is essential for developing athlete-centered educational strategies and promoting evidence-based practices in sports nutrition. This study aims to (1) determine the prevalence of dietary supplement use among Filipino national athletes and (2) examine its association with demographic and sport-related factors, including age, gender, sports discipline, training experience, education level, sources of information, and reasons for supplement use. A cross-sectional analytical design will be employed among Filipino athletes aged 18 years and above who are registered under their respective national sports associations and are free from chronic illnesses or medical treatment. Quota sampling will be utilized to ensure representation across sports disciplines. Data will be collected using the Questionnaire on the Use of Dietary Supplements adapted from Broniecka et al. (2025). Approval from the Philippine Sports Commission has been sought, and a minimum of 121 participants will be recruited to achieve 80% statistical power. The questionnaire includes supplement use patterns. Data collection may be conducted online or on-site, depending on participant availability and location, with research assistants available to clarify questions. Descriptive and inferential statistics will be analyzed using RStudio and Jamovi to examine associations between supplement use and selected variables. Ethical approval was obtained from the UST Ethics Review Committee. Findings are expected to bridge local knowledge gaps, enhance athlete health and performance, support strategic nutrition education programs, and strengthen institutional support and policy development related to athlete nutrition and supplementation practices in sports. Moreover, this study forms part of Phase One of the Philippine National Anti-Doping Organization's Project AGILA, a World Anti-Doping Agency-compliant digital ecosystem and Electronic Medical Record system designed to protect athletes from accidental doping, promote the safe and evidence-based use of dietary supplements, and support overall athlete health, readiness, and compliance with anti-doping regulations.

**Keywords:** Dietary supplement, National Athletes, Sports Nutrition, Prevalence

## References

[1] Darvishi L, Askari G, Hariri M, Bahreynian M, Ghiasvand R, Ehsani S, et al. The use of nutritional supplements among male collegiate athletes. Int J Prev Med. 2013;4(Suppl 1):S68-72.

[2] Maughan RJ, Burke LM, Dvorak J, Larson-Meyer DE, Peeling P, Phillips SM, et al. IOC consensus statement: dietary supplements and the high-performance athlete. Br J Sports Med. 2018;52(7):439-55. doi: 10.1136/bjsports-2018-099027.

[3] Garthe I, Maughan RJ. Athletes and supplements: prevalence and perspectives. Int J Sport Nutr Exerc Metab. 2018;28(2):126-38. doi: 10.1123/ijsnem.2017-0429.

[4] Knapik JJ, Steelman RA, Hoedebecke SS, Austin KG, Farina EK, Lieberman HR. Prevalence of dietary supplement use by athletes: systematic review and meta-analysis. Sports Med. 2016;46(1):103-23. doi: 10.1007/s40279-015-0387-7.

[5] Broniecka A, Sarachman A, Zagrodna A, Książek A. Dietary supplement use and knowledge among athletes: prevalence, compliance with AIS classification, and awareness of certification programs. J Int Soc Sports Nutr. 2025;22(1):2496450. doi: 10.1080/15502783.2025.2496450.

# Interprofessional collaboration competency development in healthcare students during clinical placements in the time of COVID-19: a mixed methods systematic review

[1]Roi Charles Pineda, [2]Priya Martin,[3]Kimberly Khor, [4]Jocel M. Regino, [5]Lauren Smith, [6]Romeo-Luis F. Ramirez and [7]Michael Palapal Sy

[1]Research Associate, University of Leuven, Belgium
[2]Associate Professor, University of Southern Queensland, Australia
[3]Medical Student, University of Queensland, Australia
[4]Assistant Professor, University of Santo Tomas, Philippines
[5]Research Assistant, University of Southern Queensland, Australia
[6]Instructor, De La Salle Medical and Health Sciences Institute, Philippines
[7]Senior Researcher, Zurich University of Applied Sciences, Switzerland

**ABSTRACT**

**Background**. The COVID-19 pandemic severely disrupted traditional clinical education for healthcare students. Consequently, the delivery and prioritization of interprofessional education and collaborative practice (IPECP) training shifted drastically toward remote and virtual platforms. Because of this shift, it remains unclear how these alternative placement models affected students' collaborative practice readiness.

**Objectives**. This review to identify which IPECP competencies developed in pre-qualification healthcare students while on clinical placements during COVID-19 pandemic and to identify the barriers and enablers in facilitating the development of IPECP competencies during the pandemic.

**Methods.** A mixed-methods systematic review was conducted following PRISMA guidelines. Major databases and supplementary sources were searched for primary research published from 2020 onward. Twenty studies met the inclusion criteria. Extracted data were synthesized using a convergent segregated approach and mapped against the Canadian Interprofessional Health Collaborative (CIHC) Competency Framework.

**Results**. Student perceptions indicate that these alternative placements supported the learning of interprofessional collaboration competencies. Outcomes mapped against the updated Canadian Interprofessional Health Collaborative Competency Framework indicate that the most frequently reported interprofessional collaboration competency was team communication and the least reported were collaborative leadership and team differences/disagreements processing. Although gains in interprofessional collaboration competencies were reported across the studies, their methodological shortcomings make

it difficult to determine whether alternative placements (e.g. online and telephone-based) were better or comparable to traditional placements (i.e. with face-to-face interactions), for interprofessional collaboration competency development.

**Conclusion**. Non-traditional and adapted clinical placement models successfully fostered essential IPECP competencies during the pandemic. These findings offer valuable guidance for designing resilient interprofessional education programs capable of withstanding future public health emergencies or localized disruptions.

Keywords: Clinical clerkship; COVID-19 pandemic; healthcare students; interprofessional education; telehealth

| Enablers | Barriers |
|---|---|
| Communication<br>• Effective communication (Lestari et al., 2023)<br>• Virtual platforms allowing and supporting communication and teamwork (Major et al., 2023) | COVID-19 pandemic restrictions<br>• Reduced opportunities for in-person teamwork and challenges in virtual interactions (Major et al., 2023) |
| Role clarity of team members<br>• Clear role definitions within teams (Lestari et al., 2023) | Limited exposure and practical experience<br>• Restrictions on direct patient care and limited team exposure (Castro et al., 2021; Lestari et al., 2023) |
| Curriculum integration and faculty support<br>• Longitudinal curriculum incorporating IPE (Bautista et al., 2020)<br>• Active faculty support and mentorship (Bautista et al., 2020; Leiman et al., 2021) | Technological and logistical challenges<br>• Technology issues, connectivity, and coordination challenges (Castro et al., 2021; Palmer et al., 2023; Pittman et al., 2023; Walker et al., 2022; Winship et al., 2020) |
| Virtual and hybrid formats offer flexibility<br>• Flexibility in delivery formats of learning activities afforded by virtual IPE activities (Major et al., 2023; Mozer et al., 2023; Neumann-Podczaska et al., 2021; Palmer et al., 2023; Winship et al., 2020) | Curriculum and resource constraints<br>• Traditional curricula and resource limitations (Bautista et al., 2020; Major et al., 2023; Mozer et al., 2023; Taylor et al., 2021)<br>• Need for proper training and development of staff involved in delivering IPE (DaSilva et al., 2022; Palmer et al., 2023) |
| Practical experience and feedback<br>• Hands-on clinical experiences and faculty feedback (Bautista et al., 2020; Castro et al., 2021; DaSilva et al., 2022; Mozer et al., 2023; Taylor et al., 2021) | Uncertainty of expectations<br>• Unclear role expectations and insufficient pre-placement frameworks (Kelly et al., 2023) |

**References:**

Bode, S., Dürkop, A., Wilcken, H., Peters, S., & Straub, C. (2021). Interprofessional learning during SARS-CoV-2 (COVID-19) pandemic conditions: The learning project i-recover as a substitute for a rotation on an interprofessional training ward. *GMS Journal for Medical Education*, *38*(1), 1–8. https://dx.doi.org/10.3205/zma001409

Canadian Interprofessional Health Collaborative. (2024, April). *Cihc competency framework for advancing collaboration*. CIHC Competency Framework for Advancing Collaboration

**PRESENTATION TYPE**

/ Oral Presentation

☐ Poster Presentation

**CONFERENCE TOPIC AREA**

☐ Social Inequality & Economic Disparities

/ Public Health & Well-being

☐ Education & Social Change

☐ Environmental & Sustainability Issues

☐ Technology & Society

☐ Conflict Resolution & Human Rights

☐ Governance & Policy Making

☐ Artificial Intelligence

☐ Wireless Technologies

☐ Cyber Security

☐ Sustainable Technology & Emerging Technologies

☐ Other (please specify): _____

# Creating a Community of Practice for Implementation Capacity Building: Komunidad ng Pagsasanay Para sa Pagpapaunlad ng Kapasidad sa Pagpapatupad- The KAPPA Project

**Serra, Mark Angel B.**[1*]; **Dones, Valentin C. III.**[1]; **Kent, Bridie**[2]; **Underwood, Frazer.**[2,*]

[1] Department of Physical Therapy, University of Santo Tomas, Manila, Philippines

[2] University of Plymouth, England, United Kingdom

***** *Corresponding author: mbserra@ust.edu.ph*

The KAPPA Project (Komunidad ng Pagsasanay para sa Pagpapaunlad ng Kapasidad sa Pagpapatupad) is an international collaboration between UK and Philippine institutions to strengthen the capacity of health professionals to implement evidence-based practice in health, care, and public health settings. It responds to the global need for effective support to translate evidence into practice, particularly in low- and middle-income country contexts where health inequities persist. The project aims to build capacity and capability for evidence implementation, co-create an intercontinental Community of Practice (CoP), and support progress toward JBI endorsement and affiliated group status for Philippine partner centers. It seeks to enhance participants' knowledge, confidence, and leadership in evidence-based practice while generating exemplar implementation projects aligned with local clinical priorities. A year-long program, co-led by a steering group, will adapt a tested UK model using the JBI Evidence Implementation Model and PACES. Twenty participants from Philippine institutions will receive online, "bite-sized" education, develop locally governed implementation projects, and engage in an online CoP blending structured teaching, informal networking, and mentorship from UK-based Filipino health professionals. Evaluation will use pre/post Evidence-Based Practice Questionnaire data and qualitative feedback. Expected results include a sustainable CoP, completion of JBI Evidence Implementation Training by a local trainer, exemplar projects in participating centers, and increased self-reported evidence-based practice capability. The project is expected to embed sustainable implementation expertise, deepen institutional partnerships, and

offer a scalable model to advance health equity through more consistent, contextually adapted use of best evidence in clinical practice.

**Keywords:** *Community of Practice, Evidence Implementation, Capacity Building, Health Professionals, JBI Model, Health Equity*

## References

[1] Underwood F, Kent B, Rae P, O'Keeffe K, Shawe J. *The South West Clinical School model: an initiative to transform clinical academic career pathways*. Journal of Research in Nursing. 2023;28(6-7):531-542. doi:10.1177/17449871231209037

# GENERATIVE AI VS STUDENT: A CROSS-SECTIONAL ANALYSIS OF CRITICAL APPRAISALS IN PHYSICAL THERAPY RESEARCH

L.P.D. Tangcuangco*[1,2], C.J.T. Escuadra[1,2], A.S.P. Caballero, E.A.P. Ferrer[1], J.A.D. Manalo[1], S.A.U. Manubay[1], A.G.M. Rabacal[1], Z.R. Teodosio[1]

*Department of Physical Therapy, College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*[1]

*Center for Health Research and Movement Science, College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*[2]

**Corresponding author:* ldtangcuangco@ust.edu.ph

Critical appraisal is essential for evidence-based practice (EBP) in physical therapy (PT) but faces barriers like time constraints and inconsistent evaluations. Generative AI tools, such as ChatGPT, present an innovative solution by offering rapid and structured appraisals. This study assessed the quality of critical appraisals generated by ChatGPT on PT research by comparing them to the rating scores of critical appraisals conducted by undergraduate PT students. A quantitative, analytic, cross-sectional design was utilized to compare retrospective data from anonymized student appraisals and AI-generated outputs. Using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Randomized Controlled Trials, 27 journal articles were purposively sampled. Descriptive statistics and the Mann-Whitney U Test ($p < 0.05$) were employed to analyze appraisal score differences. Results revealed no significant difference in total appraisal scores between AI and student groups ($p=0.48$). The two groups demonstrated high agreement (74.10%–92.60%) on 11 of 13 JBI items, including randomization, blinding, and statistical analysis. However, significant differences emerged in items requiring mechanistic reasoning and contextual interpretation, specifically the concealment of group allocation ($p=0.01$) and similarity of identical treatment procedures ($p=0.01$). For these items, AI responses exhibited higher variability and tended to evaluate items as "unclear" when explicit statements were absent, whereas human raters deduced context more effectively. ChatGPT's critical appraisal capabilities are comparable to those of PT students for straightforward methodological criteria. However, AI exhibits limitations in interpreting implicit details requiring human judgment. A human-AI collaborative approach is recommended for EBP education, utilizing AI as a supportive preliminary scaffold while relying on human oversight for complex methodological evaluations.

**Keywords:** ChatGPT; physical therapy research; critical appraisal; evidence-based practice; artificial intelligence

| Questions | Similar rating (n,%) | Different rating (n,%) | p-value | Rank Biserial Correlation Effect Size |
|---|---|---|---|---|
| 1 - Was true randomization used for assignment of participants to treatment groups? | 23 (85.20) | 4 (14.80) | 0.064 | -0.173 |
| **2 - Was allocation to treatment groups concealed?** | **15 (55.60)** | **12 (44.40)** | **0.010** | **-0.384** |
| 3 - Were treatment groups similar at the baseline? | 26 (96.3) | 1 (3.70) | 0.336 | -0.037 |
| 4 - Were participants blind to treatment assignment? | 21 (77.80) | 6 (22.20) | 0.844 | 0.025 |
| 5 - Were those delivering treatment blind to treatment assignment? | 20 (74.10) | 7 (25.90) | 0.835 | -0.030 |
| 6 - Were outcomes assessors blind to treatment assignment? | 21 (77.80) | 6 (22.20) | 0.072 | -0.239 |
| **7 - Were treatment groups treated identically other than the intervention of interest?** | **18 (66.70)** | **9 (33.30)** | **0.009** | **-0.340** |
| 8 - Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed? | 24 (88.90) | 3 (11.10) | 0.586 | 0.036 |
| 9 - Were participants analyzed in the groups to which they were randomized? | 25 (92.60) | 2 (7.40) | 0.161 | -0.074 |
| 10 - Were outcomes measured in the same way for treatment groups? | 24 (88.90) | 3 (11.10) | 0.571 | 0.037 |
| 11- Were outcomes measured in a reliable way? | 24 (88.90) | 3 (11.10) | 0.089 | 0.108 |
| 12 - Was appropriate statistical analysis used? | 25 (92.60) | 2 (7.40) | 0.331 | -0.071 |
| 13 - Was the trial design appropriate, and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial? | 24 (88.90) | 3 (11.10) | 0.571 | -0.037 |

Table 1: Frequency and proportion of similar and different ratings given by AI and student raters

References:

1. Sackett, D. L. (1997). Evidence-based medicine. Semin Perinatol, 21(1), 3-5. https://doi.org/10.1016/s0146-0005(97)80013-4
2. The Joanna Briggs Institute. (2017). Checklist for Randomized Controlled Trials. The Joanna Briggs Institute.
3. Tarakji, Z., Kanaan, A., Saadi, S., et al. (2024). Concordance between humans and GPT-4 in appraising the methodological quality of case reports and case series using the Murad tool. BMC Med Res Methodol, 24(1), 266. https://doi.org/10.1186/s12874-024-02372-6
4. Hao, J., Yao, Z., Tang, Y., et al. (2025). Artificial Intelligence in Physical Therapy: Evaluating ChatGPT's Role in Clinical Decision Support for Musculoskeletal Care. Ann Biomed Eng, 53, 9-13. https://doi.org/10.1007/s10439-025-03676-4
5. Zhang, Q., & Rapport, M. J. (2025). Purposeful Integration of Artificial Intelligence in Evidence-Based Practice Course for Doctor of Physical Therapy students. International Journal of Allied Health Sciences and Practice, 23(2).

**PRESENTATION TYPE**

☐ Oral Presentation
✔ Poster Presentation

**CONFERENCE TOPIC AREA**

☐ Social Inequality & Economic Disparities
☐ Public Health & Well-being
✔ Education & Social Change
☐ Environmental & Sustainability Issues
☐ Technology & Society
☐ Conflict Resolution & Human Rights
☐ Governance & Policy Making
☐ Artificial Intelligence
☐ Wireless Technologies
☐ Cyber Security
☐ Sustainable Technology & Emerging Technologies
☐ Other (please specify): _____

# Exploring Teachers' Perspectives on Facilitating Social Skills Among Learners with Special Needs in Inclusive Schools in Quezon City: A Qualitative Descriptive Study

**Bautista, Elinor C.; Faminial, Shyruel Rose M.; Cabradilla, Kirsten Carys A.; Mamitag, Shen Mee Coleen L.; Mangahas, Marielle Anne; Sindayen, Patrick D.; Ymballa, Jianna B.**

[1]*Department of Speech-Language Pathology, University of Santo Tomas, Manila, Philippines*
[2]*College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*
** Corresponding author: elcbautista@ust.edu.ph*

Social skills play a vital role in a student's academic performance and overall development. Although speech-language pathologists directly target social skills, teachers are considered critical agents in their ongoing facilitation, as they spend most of their time in the classroom. This study aims to describe the perspectives of teachers on the strategies to support social skills and the factors that hinder their facilitation. A qualitative descriptive design was employed where nine public and private inclusive education teachers were recruited through purposive sampling. Data was collected through face-to-face and online interviews using a semi-structured interview guide. All data were analyzed using Braun and Clarke's six-step thematic analysis framework, with NVivo software used to aid the analysis. Three themes emerged from the study: (1) Teachers' Notions of Social Skills, (2) Teachers Create Interactions in Inclusive Classrooms, and (3) Teachers Recognize the Multifaceted Contextual Factors to Social Inclusion. The findings of the study show that social skills are defined as essential competencies that enable learners to interact and adapt to social contexts, making them crucial for long-term development. Teachers also view themselves as facilitators of social skills through individualized instruction, providing opportunities for peer interaction, fostering peer-mediated support, and interprofessional collaboration. However, barriers such as limited knowledge and training, persistent stigma, and lack of family support hindered effective facilitation. In conclusion, teachers' efforts to facilitate social skills are shaped by contextual factors, often creating a gap between their intentions and capacity. Stronger systemic support, collaboration, and sustained professional development are required for effective facilitation.

**Keywords:** social skills, inclusive schools, learners with special needs

**PRESENTATION TYPE**
- ☐ Oral Presentation
- ✔ Poster Presentation

**CONFERENCE TOPIC AREA**
- ☐ Social Inequality & Economic Disparities
- ☐ Public Health & Well-being
- ✔ Education & Social Change
- ☐ Environmental & Sustainability Issues
- ☐ Technology & Society
- ☐ Conflict Resolution & Human Rights
- ☐ Governance & Policy Making
- ☐ Artificial Intelligence
- ☐ Wireless Technologies
- ☐ Cyber Security
- ☐ Sustainable Technology & Emerging Technologies
- ☐ Other (please specify): _____

# Exploring Interventions and Strategies Used by School-Based Speech-Language Pathologists for Grade School Students within the National Capital Region of the Philippines: A Qualitative Descriptive Study

**Blanco, Pia Angela, J.; Bautista, Elinor C.; Agbay, Czarine Aziza F.; Gamo, Marcus Gerome D.; Javier, Jorjana Betina N.; Santiago, Allyssa Fayth R.; Sison, Carl Angela M.; Weir, Krizzia Cybelle P.**
*Department of Speech-Language Pathology, University of Santo Tomas, Manila, Philippines*
*College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*
** Corresponding authors: elcbautista@ust.edu.ph, pjblanco@ust.edu.ph*

Communication disorders may create long-term barriers that can impact children, leading to difficulties catching up academically and socially [1]. In school settings, Speech-language Pathologists (SLPs) play an essential role in supporting students' communication, literacy, and participation through individualized and collaborative interventions [2]. Despite the growing recognition of school-based SLPs in the Philippines, limited local research has explored their intervention practices and decision-making processes [3]. This qualitative descriptive study explored the interventions, strategies, and contextual factors influencing the practices of school-based SLPs working with grade school students within the National Capital Region (NCR) of the Philippines. Ten licensed school-based SLPs with at least two years of experience participated in semi-structured one-on-one interviews conducted through Zoom. Data gathered were transcribed verbatim and analyzed using thematic analysis. Findings revealed that school-based SLPs implement diverse intervention approaches depending on the communication and academic needs of students. Language-focused interventions, literacy-based supports, speech sound motor-based approaches, cognitive and academic supports, behavioral strategies, and sensory-based approaches were commonly utilized by school-based SLPs working in the NCR. These are delivered through push-in and pull-out models, individual, dyad, and group-based sessions, which depend on students' readiness, classroom demands, and institutional factors. These choices were shaped by theoretical knowledge, ecologically valid assessment practices, collaboration with teachers and other professionals, and the individual strengths and needs of students. However, contextual constraints such as scheduling conflicts, large caseloads, differing professional perspectives, and limited institutional support affected intervention implementation. Overall, the findings highlight how school-based SLPs in the NCR adopt flexible, holistic, and context-responsive intervention practices to support students' communication and academic participation. The study underscores the need for localized research, stronger institutional support, and more contextually responsive school-based speech-language pathology practices in the Philippines.

**Keywords:** school-based SLPs, strategies, interventions, Qualitative Descriptive, NCR

## References

[1] Riad, R., Allodi, M. W., Siljehag, E., & Bölte, S. (2023). Language skills and well-being in early childhood education. *Frontiers in Education, 8*, 963180.

[2] Steele, S. C. (2020). Vocabulary intervention: A national survey. *Communication Disorders Quarterly, 41*(3), 151–161.

[3] Denman, D., Cordier, R., Kim, J. H., Munro, N., & Speyer, R. (2021). What influences speech-language pathologists' use of different types of language assessments for elementary school-age children? *Language, Speech, and Hearing Services in Schools, 52*(3), 776–793.

# Perspectives of Occupational Therapy Clinicians on the Use of Artificial Intelligence in Clinical Practice in Metro Manila: A Qualitative Study

**Gomez, Ivan Neil B.**[1]; **Cua, Justine Anne G.**[1]; **Endozo, Catherina Moira A.**[1,*]; **Berja, Kariza Gale F.**[1]; **Go, Gabriel Derick C.**[1]; **Lagman, Sabina Diorela Simone G.**[1]; **Rodriguez, Jenny Lynn S.**[1]

[1] Department of Occupational Therapy, University of Santo Tomas, Manila, Philippines

***** *Corresponding author: catherinamoira.endozo.crs@ust.edu.ph*

Artificial intelligence (AI) is increasingly transforming healthcare, including its emerging applications in occupational therapy (OT), with generative artificial intelligence (GenAI) showing potential to enhance clinical practice [1]-[3]. However, its integration depends not only on technological advancements but also on clinicians' perceptions, acceptance, and experiences, which remain underexplored [3]. This qualitative phenomenological study aimed to explore the perspectives of OT clinicians in Metro Manila, Philippines on the use of AI, particularly GenAI, in clinical practice, focusing on their experiences, attitudes, acceptance, and intention to use the technology. Guided by the Technology Acceptance Model (TAM), the study involved 16 OT clinicians working across Metro Manila who were familiar with AI and with at least one year of clinical experience, excluding those in academic roles. Participants were also purposively sampled to ensure variation in fields of practice and clinical settings. Following a successful pilot test, one-on-one semi-structured interviews were conducted online and thematically analyzed using NVivo 15 based on a coding framework. Five themes emerged: (i) AI as a supportive but bounded clinical tool, (ii) safeguarding professional identity and clinical judgment, (iii) differences in user competence shaping effective AI use, (iv) conditional acceptance of AI based on perceived risks, and (v) context-informed recommendations for AI use and improvement. Participants generally viewed AI as useful for improving efficiency in clinical and administrative tasks, but expressed concerns regarding reliability, ethics, privacy, and contextual resource constraints. Overall, acceptance of AI was generally positive but cautious, with its use seen as highly context-dependent and requiring professional judgment. The findings highlight the need for training, clear guidelines, and responsible integration to support effective and ethical use in OT practice.

**Keywords:** artificial intelligence; occupational therapy; perspectives; perceptions; qualitative

**Table 1.** Summary of Themes and Sub-themes Generated in this Study

| Theme | Sub-Themes |
|---|---|
| (1) AI as a Supportive but Bounded Clinical Tool | 1.1 AI as a supportive adjunct tool in practice<br>1.2 AI as an idea generator and efficiency enhancer |
| (2) Safeguarding Professional Identity and Clinical Judgment | 2.1 Preserving clinical judgment in AI-assisted practice<br>2.2. Concerns with the reliability and regulation of AI |
| (3) Differences in User Competence Shape Effective AI Use | 3.1 Varying knowledge and skills with emerging AI<br>3.2 Understanding AI as an adaptive yet user-dependent tool |
| (4) Conditional Acceptance of AI Based on Perceived Risks | 4.1 Receptive yet guarded views on AI use<br>4.2 Resistance shaped by perceived risks of AI |
| (5) Context-Informed Recommendations for AI Use and Improvement | 5.1 Contextual opportunities and constraints in AI use<br>5.2 Advancing systems and capacities for responsible AI use |

## References

[1] Wang, F. & Preininger, A. (2019). AI in Health: State of the Art, Challenges, and Future Directions. *Yearbook of Medical Informatics*, 28(1), 16–26. https://doi.org/10.1055/s-0039-1677908.

[2] Hirani, R., Noruzi, K., Khuram, H., Hussaini, A., Aifuwa, E.I., Ely, K., et al., (2024). Artificial Intelligence and Healthcare: A Journey through History, Present Innovations, and Future Possibilities. Life 14(5), 557. https://doi.org/10.3390/life14050557.

[3] Nah, F.F.-H., Zheng, R., Cai, J., Siau K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, 25(3), 277-304. https://doi.org/10.1080/15228053.2023.2233814

# Outcomes Insights: A Learning Analytics Dashboard for Entry-Level Physical Therapy Learning Outcomes in a University in the Philippines

C.J.T. Escuadra[1,2]

[1]Department of Physical Therapy, College of Rehabilitation Sciences, University of Santo Tomas
[2]Department of Education, Ewha Womans University
*Corresponding author: ctescuadra@ust.edu.ph*

The use of a learning analytics dashboard (LAD) is a systematic way to analyze and monitor program-intended learning outcomes and to provide feedback to help improve students' learning strategies. Giving students the opportunity to view and understand their status in such an application allows them to be more involved and empowered in their learning process. Past studies in Namoun and Alshanqiti's (2020) review have highlighted that each institution should create its own learning analytics system to account for the specific characteristics of its students and curriculum. This study aims to develop and test a prototype LAD called Outcomes Insight and to determine its effectiveness in the learning and study strategies of selected Filipino university students. This research study had four sequential phases following an iterative design-based research described by Wang and Hannafin (2015) and Yoo & Jin (2020). In phase 1, predictive models were developed using structural equation modeling and artificial neural network (SEM-ANN) on the final grades of 399 physical therapy students. In phase 2, Outcomes Insight was developed using the predictive models and evaluated by three education experts. In phases 3-4, the acceptability and effects of the dashboard on the learning and study strategies of 28 second-year physical therapy students were tested using the Learning and Study Strategies Inventory (LASSI) 3rd edition. Data from phases 3 and 4 were analyzed using descriptive and inferential statistics. For phase 1, relevant courses for predicting each program outcome were identified with acceptable fit metrics in SEM-ANN (CFI & TLI>0.90, RMSEA & SRMR<0.08, MAE & MSE=0.071-0.418). For phase 2, the developed prototype included three distinct but related pages: input, results, and feedback pages. Experts and students agreed that the dashboard includes relevant elements and a visually appealing, accessible, and easy-to-use structure. For Phase 4, significant changes ($p<0.05$) and medium to large effect sizes (0.520-0.881) were found across students' habits, professional development, and mind. The developed and validated Outcomes Insight prototype presents an excellent potential for providing descriptive, predictive, and prescriptive feedback to improve physical therapy students' learning and study strategies. The dashboard included comprehensive, accurate predictive models and personalized recommendations to help students understand their current status and potential improvement options. Health educators should continuously use innovative strategies and relevant data to monitor students' progress and provide personalized recommendations.

Keywords: Learning analytics dashboard, Structural Equation Modeling, Artificial Neural Network

**Table 1.** *Summary of LASSI Scores of students at pre and post-tests*

| (Total=28) | *Pre-Test* | | *Post-Test* | | *Statistic* | *p-value* | *Effect Size* | Interpretation |
|---|---|---|---|---|---|---|---|---|
| | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Habits of Professionals | | | | | | | | |
| Attitude | 43.10 | 23.50 | 54.00 | 23.90 | 78.0 | 0.023 | 0.520 | Moderate |
| Motivation | 44.30 | 27.10 | 52.80 | 26.40 | 84.0 | 0.020 | 0.521 | Moderate |
| Time Management | 49.10 | 24.0 | 50.60 | 23.10 | 167.5 | 0.848 | 0.046 | - |
| Average | 44.90 | 20.50 | 52.50 | 18.80 | 64.5 | 0.013 | 0.659 | Large |
| Habits of Learning | | | | | | | | |
| Information Processing | 74.40 | 19.50 | 74.30 | 21.50 | 106.0 | 0.672 | 0.116 | - |
| Selecting Main Ideas | 37.10 | 25.90 | 47.00 | 26.90 | 73.0 | 0.049 | 0.470 | Moderate |
| Self-Testing | 75.30 | 19.90 | 81.20 | 13.40 | 80.0 | 0.075 | 0.420 | - |
| Test Strategies | 40.00 | 23.50 | 45.20 | 24.20 | 118.5 | 0.239 | 0.271 | - |
| Using Academic Resources | 35.50 | 24.80 | 51.60 | 23.40 | 12.5 | <0.001 | 0.881 | Large |
| Average | 52.40 | 15.00 | 59.90 | 12.60 | 41.0 | <0.001 | 0.766 | Large |
| Habits of Mind | | | | | | | | |
| Anxiety | 20.60 | 19.90 | 35.50 | 23.50 | 34.5 | 0.002 | 0.750 | Large |
| Concentration | 36.60 | 28.30 | 44.00 | 24.00 | 50.5 | 0.075 | 0.468 | - |
| Average | 28.50 | 19.50 | 39.70 | 19.50 | 33.5 | <0.001 | 0.809 | Large |

## References

[1] Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), 237.

[2] Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. Educational technology research and development, 53(4), 5-23.

[3] Yoo, M., & Jin, S. H. (2020). Development and evaluation of learning analytics dashboards to support online discussion activities. Educational Technology & Society, 23(2), 1-18.

**PRESENTATION TYPE**

☐ Oral Presentation

■ Poster Presentation

**CONFERENCE TOPIC AREA**

☐ Social Inequality & Economic Disparities
☐ Public Health & Well-being
☐ Education & Social Change
☐ Environmental & Sustainability Issues
☐ Technology & Society
☐ Conflict Resolution & Human Rights
☐ Governance & Policy Making
☐ Artificial Intelligence
☐ Wireless Technologies
☐ Cyber Security
■ Sustainable Technology & Emerging Technologies
☐ Other (please specify): _____

# Characterizing Musical Parameters for Stress Intervention: A Biofeedback-Based Analysis of Physiological and Neurophysiological Responses

**Fabay, Lianna Mae A.**[1,2,*]; **Ocenar, Samuel John A.**[1,2]; **Danganan, Lester Mico L.**[5]; **Nieles, Jay Patrick M.**[1]; **Catindig-Dela Rosa, Joseree Ann S.**[6,7]; **Prado, Seigfred V.**[1,2,3,4]

[1] Brain Research Advancement and Innovations in Neurotechnology (BRAIN) Lab, University of Santo Tomas, Manila, Philippines
[2] Graduate School, University of Santo Tomas, Manila, Philippines
[3] Department of Electronics Engineering, University of Santo Tomas, Manila, Philippines
[4] Research Center for Engineering and Technological Sciences, University of Santo Tomas, Manila, Philippines
[5] NSW Health, Sydney, New South Wales, Australia
[6] Faculty of Medicine and Surgery, University of Santo Tomas, Manila, Philippines
[7] National Center for Geriatric Health, Manila, Philippines
***** *Corresponding author: liannamae.fabay.gs@ust.edu.ph*

Stress is a psycho-physiological response to overwhelming demands, often triggered by academic pressures, and is most prevalent among college students, reporting the highest levels of perceived stress. Pharmacological coping strategies, such as the use of stimulants and sedatives, are widely practiced but pose risks of dependency, cognitive fatigue, and limited long-term effectiveness. As a result, safer and more sustainable alternatives are necessary. Among the available non-pharmacological approaches, music therapy (MT) has consistently demonstrated superior efficacy in reducing stress and anxiety due to its influence on both emotional regulation and autonomic control. While MT has shown promising results in improving mood and lowering physiological arousal, most current studies rely on subjective reporting and lack objective neurophysiological evaluation. This study developed a quantitative framework to evaluate the effects of musical parameter combinations on EEG-based neural responses and on smartwatch-monitored physiological responses during a music-based stress intervention. Musical stimuli were selected from the DEAM dataset using a feature-based modeling approach, in which extracted audio features were reduced via Principal Component Analysis and grouped using *K*-means clustering, yielding 12 representative music excerpts. Neurophysiological data were recorded using an 8-channel OpenBCI EEG system, while physiological markers, including heart rate, electrodermal activity, and skin temperature, were captured using an EmotiBit physiological monitoring smartwatch. Lab Streaming Layer (LSL) was used to synchronize multimodal signals across baseline, stress, music listening, and post-intervention

phases. EEG and physiological data were processed to extract stress-relevant features and were prepared for analysis using Canonical Correlation Analysis (CCA) and Linear Mixed-Effects Modeling (LMM). Through this integrative approach, the study provided an objective basis for identifying music parameters that support personalized, scalable, and non-invasive stress intervention.

**Keywords:** Music Therapy, Stress Intervention, EEG, Physiological Monitoring, Multimodal signals

## References

[1] Barbayannis G, Bandari M, Zheng X, Baquerizo H, Pecor KW, Ming X. Academic Stress and Mental Well Being in College Students: Correlations, Affected Groups, and COVID 19. *Frontiers in Psychology*. 2022;13:886344. doi: 10.3389/fpsyg.2022.886344.

[2] Fedotchev, A., Parin, S., Polevaya, S., & Zemlianaia, A. (2022). EEG based musical neurointerfaces in the correction of stress induced states. *Brain Computer Interfaces, 9*(1), 1 to 6.

[3] Amanvermez, Y., Rahmadiana, M., Karyotaki, E., De Wit, L., Ebert, D. D., Kessler, R. C., & Cuijpers, P. (2020). Stress management interventions for college students: A systematic review and meta analysis. *Clinical Psychology Science and Practice, 30*(4), 423 to 444. https://doi.org/10.1111/cpsp.12342

[4] Paszkiel S, Dobrakowski P, Łysiak A. The Impact of Different Sounds on Stress Level in the Context of EEG, Cardiac Measures and Subjective Stress Level: A Pilot Study. *Brain Sciences*. 2020;10(10):728. doi: 10.3390/brainsci10100728.

[5] El Makrini, I., Turcksin, T., Incirci, T., Thiery, E., Kindt, S., Lovecchio, R., Cao, H. L., Denayer, M., Lamine, E., Huysentruyt, S., Verstraten, T., & Vanderborght, B. (in press). PhysioSense: An Open Source Multi Modal Monitoring Framework for Human Movement and Behavior Analysis. *IEEE Robotics & Automation Magazine*.

# Perspectives and Experiences of Physical Therapy Interns on Telerehabilitation in the Philippines

**Fernandez, Ma. Roxanne L.[1],[2]*; Cruz, Christopher G.[1]; Co, Mary Therese A.[1]; Aypa, Nicolas Macwain G.[1]; Dy-Lim, Vanna Chelsey C.[1]; Molina, Philip Louie L.[1]; Pasco, Julianne Kimberly U.[1]; Relucio, Russ Jason L.[1]; Ugbinar, Asia C.[1]**

[1] Department of Physical Therapy, University of Santo Tomas, Manila, Philippines
[2] Center for Health Research and Movement Science, University of Santo Tomas, Manila, Philippines
* *Corresponding author: mlfernandez@ust.edu.ph*

The COVID-19 pandemic accelerated the integration of telerehabilitation into physical therapy (PT) education and practice, aiming to maintain accessibility, affordability, and safety in patient care. By using tools such as video conferencing and remote exercise monitoring, telerehabilitation supported clinical learning during rotations. Understanding the perspectives and experiences of PT interns offers important insights into their role in patient care and future curriculum development. This study explored the perspectives and experiences of PT interns regarding telerehabilitation, focusing on preparedness, challenges, adaptive strategies, and perceptions of clinical effectiveness. A descriptive qualitative design guided by the Normalisation Process Theory (NPT) and the Unified Theory of Acceptance and Use of Technology (UTAUT). Participants were recruited through purposive sampling. While recruitment targeted PT graduates nationwide who had prior telerehab experience, participants who met the eligibility criteria came from Luzon and Iloilo in the Visayas region. Three online focus group discussions were conducted using a researcher-developed interview guide. Sessions were recorded, transcribed verbatim, and analysed using Braun and Clarke's six-step thematic analysis. Thematic analysis generated several key themes: Adapting to a New Mode of Practice, Logistical Barriers, Documentation and Assessment Demands, Home-Based Relevance, Insufficient Preparation, Competence through Experience, Supervisory Support, Technological Familiarity, Limitations of Remote Assessments, Developing Professional Composure, Confronting Contextual and Technical Barriers, Increased Reach, Early Doubts, Growing Engagement, Influence of Caregiver Support, Perceived Effectiveness of Telerehabilitation, Impact of Realistic Goal-Setting, Evaluation of Technologies and Tools Used, Technological Barriers, Sustainability and Accessibility, Educational Integration, and Professional Disposition. In conclusion, telerehabilitation was perceived as a meaningful yet demanding form of care, highlighting the need for structured training, clear protocols, and adequate supervision to support its effective integration into PT education and practice.

**Keywords:** telerehabilitation; physical therapy education; clinical education; qualitative research; normalisation process theory; unified theory of acceptance and use of technology

# Feature Characterization using IMU Kinematic Data to Aid in Diagnosing Patellofemoral Pain Syndrome

***Xhan Benedict Grefaldeo[1], Jade Marco L. Idio[1], Josiah David L. La Guardia[1], Caesar Manuel S. Marquez[1], Consuelo B. Gonzales-Suarez[2,3], Ivan Neil B. Gomez[4], Jan-Tyrone C. Cabrera[2,3], Jazzmine Gail S. Flores[3], Edison A. Roxas[1], Mary Grace Ann C. Bautista[1], Emmanuel C. Guevara[1], Gabriel Rodnei M. Geslani[1]**

*[1] Department of Electronics Engineering, University of Santo Tomas, Manila, Philippines*
*[2] Department of Physical Medicine and Rehabilitation, University of Santo Tomas, Manila, Philippines*
*[3] Research Center for the Health Sciences, University of Santo Tomas, Manila, Philippines*
*[4] Department of Occupational Therapy, University of Santo Tomas, Manila, Philippines*
**Corresponding author: gmgeslani@ust.edu.ph*

This study addresses the diagnostic limitations of binary Patellofemoral Pain Syndrome (PFPS) classification by utilizing Inertial Measurement Unit (IMU) data to characterize PFPS-indicative kinematic features. Principal Component Analysis (PCA) was applied to high-dimensional movement data, retaining 16 components that explained 90.17% of the total variance. These components were evaluated using filter, embedded, and wrapper feature selection methods to identify those most relevant for PFPS classification. Preliminary findings show that the wrapper method achieved the highest balanced accuracy of 76% using two principal components (PC9 and PC16), while all selection frameworks consistently converged on two biomechanical domains: hip-femur rotational mechanics and lumbar trunk coordination. Key features included internal hip and femur rotation and increased lumbar rotational excursion. These results provide proof-of-concept evidence that PFPS is better characterized by coordinated multi-segment interactions rather than isolated joint abnormalities. The findings suggest strong generalization potential for identifying biomechanical signatures of PFPS, supporting future work toward improved clinical diagnosis and personalized treatment strategies.

# Satisfaction of UST-CRS Students with the Quality of Blended Teaching and Learning in the Interprofessional Education Course for Academic Year 2025-2026: A Cross-Sectional Study

**Dones, Valentin C. III[1]; Versales, Cristine Rose S.[1]; Alvaro, Donita Faye D.[1]; Carpio, John Daniel D.[1]; Garces, Maiquela V.[1]; Ledesma, Antonius Andrei L.[1*]; Sison, Juan Antonio M.[1]; Tabios, Kate Riane S.[1]**

[1] Department of Physical Therapy, University of Santo Tomas, Manila, Philippines

***** *Corresponding author: andreiledesma2@gmail.com*

Blended teaching and learning (BTL) has become increasingly integrated into higher education settings as it combines face-to-face and online learning modalities in order to ensure greater flexibility, engagement, and accessibility. In the Philippines, BTL is aligned with the Commission on Higher Education guidelines. However, local evidence regarding its implementation in health-allied education remains limited. This study evaluated the quality of BTL implementation in the Interprofessional Education (IPE) course of the University of Santo Tomas College of Rehabilitation Sciences (UST-CRS) for Academic Year 2025-2026 by assessing the satisfaction of third-year rehabilitation science students. For this study, a descriptive observational cross-sectional design was used and all eligible third-year UST-CRS students enrolled in the IPE course were invited to participate. Data was collected online via Google Forms using the Blended Teaching Assessment Tool (BTAT), a 37-item questionnaire that measures satisfaction across six dimensions: Course Direction and Guide, Quality Teaching Delivery Skills, Quality of Instructional Media and Recorded Videos, Use of Educational Technology and Instructional Media, Quality of Learning Activities, and Reflective Feedback. A total of 102 students completed the questionnaire, representing a 44% response and completion rate. Descriptive statistics, including mean scores, standard deviations, and 95% confidence intervals, were used to analyze the data gathered. Results showed generally favorable satisfaction across all six BTAT dimensions, with mean scores ranging from 3.075 to 3.561. Dimension III, Quality of Instructional Media and Recorded Videos, received the highest mean score, suggesting strong satisfaction with the accessibility and quality of learning materials. Conversely, Dimension VI, Reflective Feedback, obtained the lowest mean score and highest variability, indicating inconsistencies in the frequency and quality of feedback provided to students. The findings suggest that BTL implementation in the IPE course is generally satisfactory and effective in supporting student learning. However, improvements in standardized and timely reflective feedback may further enhance course delivery, student motivation, and curriculum quality in health-allied education.

**Keywords:** blended learning; student satisfaction; health-allied education; Philippines

**PRESENTATION TYPE**

☐ Oral Presentation

√ Poster Presentation

**CONFERENCE TOPIC AREA**

☐ Social Inequality & Economic Disparities

☐ Public Health & Well-being

☐ Education & Social Change

☐ Environmental & Sustainability Issues

☐ Technology & Society

☐ Conflict Resolution & Human Rights

☐ Governance & Policy Making

☐ Artificial Intelligence

☐ Wireless Technologies

☐ Cyber Security

☐ Sustainable Technology & Emerging Technologies

√ Other (please specify): Sports Psychology and Athletic Development

# Psychological Profiling of Mental Toughness, Motivation, and Grit in UST Men's and Women's Collegiate Basketball Players During the Off-Season: A Descriptive Correlational Study

R.D. Ramos[1], L.R. Cosep[1], E.N. De Mesa[1], J. Dogillo[1], A. See[1], K. Archog[1], E.Coronel[1], J. Lazo[1], H. Perez[1]

[1] *Department of Sports Science, University of Santo Tomas, Manila, Philippines*

**Corresponding author: rdramos@ust.edu.ph*

Psychological attributes such as mental toughness, motivation, and grit are fundamental in shaping how athletes respond to training demands, cope with pressure, and sustain long-term performance. While these constructs have been extensively examined during highly competitive in-season periods, limited research has explored their expression during the off-season – a critical developmental phase characterized by reduced external reinforcement, lower immediate pressure, and a greater reliance on self-regulation and internal drive [1]. Furthermore, profiling these specific psychological traits among Southeast Asian collegiate populations remains underexplored. To address this gap, this study aimed to systematically describe the psychological profile of the University of Santo Tomas (UST) Men's and Women's Collegiate Basketball Teams during the off-season and to examine the underlying relationships among mental toughness, sport motivation, and grit. A descriptive correlational cross-sectional research design was employed, utilizing total enumeration sampling to recruit 22 eligible collegiate basketball players. Participants completed three validated self-report measures: the Sports Mental Toughness Questionnaire (SMTQ-14) [2], the Sport Motivation Scale-II (SMS-II) [3], and the Domain-Specific Grit Scale for College Athletic Students (DGSCAS) [4]. Descriptive statistics were used as a primary analytical tool to summarize and quantify the baseline psychological profiles of the sample, providing a detailed overview of the athletes' psychological variables across specific subscales. The descriptive analysis revealed that athletes maintained moderate to high levels across all main constructs. Specifically, mental toughness (M = 2.96 ± 0.38) showed higher mean scores in constancy (M = 3.32 ± 0.57) and confidence (M = 3.20 ± 0.54) than in emotional control (M = 2.23 ± 0.56), suggesting a potential area for off-season psychological development. Motivation (M = 5.33 ± 0.74) was predominantly driven by self-determined forms, including identified (M = 6.06 ± 1.36), integrated (M = 5.89 ± 1.38), and intrinsic regulation (M = 5.82 ± 1.54). Grit (M = 4.00 ± 0.75)

demonstrated strong perseverance of effort (M = 4.10 ± 0.82) and consistency (M = 3.90 ± 0.82) despite the lack of competitive stimuli. Pearson correlation analysis revealed that mental toughness was significantly and positively correlated with both motivation (r = 0.483, $p$ = 0.023) and grit (r = 0.579, $p$ = 0.005), stating that athletes with higher levels of mental toughness also tend to exhibit higher levels of motivation and grit. Conversely, no statistically significant relationship was observed between motivation and grit (r = 0.282, $p$ = 0.204), stating that these constructs operate independently in this context and that the presence of one attribute does not necessarily correspond to higher levels of the other. The findings suggest that mental toughness serves as a central psychological anchor associated with both motivational processes and sustained effort. These baseline descriptive and correlational insights can critically inform coaches and sports psychologists in designing targeted mental conditioning strategies, particularly interventions aimed at enhancing emotional control and sustaining intrinsic drive outside of the competitive season.

**Keywords:** mental toughness, sport motivation, grit, off-season, collegiate basketball

**Table 1.** Descriptive Statistics and Normality of Study Variables (N = 22) & Relationship Among Main Constructs

| Variables | Level | Variable | Post-Intervention (M ± SD) | Variables | r | *p-value* |
|---|---|---|---|---|---|---|
| Mental Toughness | Overall | SMTQ Total | 2.96 ± 0.38 | Mental Toughness to Motivation | .483 | 0.023 |
| | Subscale | Confidence | 3.20 ± 0.54 | | | |
| | | Constancy | 3.32 ± 0.57 | | | |
| | | Control | 2.23 ± 0.56 | | | |
| Motivation | Overall | SMS Total | 5.33 ± 0.74 | Mental Toughness to Grit | .579 | .005 |
| | Subscale | Intrinsic | 5.82 ± 1.54 | | | |
| | | Integrated | 5.89 ± 1.38 | | | |
| | | Identified | 6.06 ± 1.36 | | | |
| | | Amotivation | 5.06 ± 1.40 | | | |
| Grit | Overall | DGSCAS Total | 4.00 ± 0.75 | Motivation to Grit | .282 | .204 |
| | Subscale | Consistency | 3.90 ± 0.82 | | | |
| | | Perseverance | 4.10 ± 0.82 | | | |

## References

[1] Chou, C., Wang, F., Wu, H., Tsai, S., Chen, C., Bernard, J. R., Kuo, Y., & Liao, Y. (2021). "The Competitive Season and Off-Season": Preliminary Research concerning the Sport-Specific Performance, Stress, and Sleep in Elite Male Adolescent Basketball Athletes. *International Journal of Environmental Research and Public Health, 18*(24), 13259.

[2] Sheard, M., Golby, J., & Van Wersch, A. (2009). Progress toward Construct Validation of the Sports Mental Toughness Questionnaire (SMTQ). *European Journal of Psychological Assessment, 25*(3), 186-193.

[3] Pelletier, L. G., Rocchi, M. A., Vallerand, R. J., Deci, E. L., & Ryan, R. M. (2012). Validation of the revised sport motivation scale (SMS-II). *Psychology of Sport and Exercise, 14*(3), 329-341.

[4] Gao, F., Wei, Q., Dong, X., Gao, J., Lu, S., & Liu, Y. (2024). Developing and validating a Domain-specific Grit Scale for College Athletic Students. *Scientific Reports, 14*(1), 11897.

# Experiences of rehabilitation professionals in palliative care: a qualitative systematic review

**Nikka Karla R. Santos[1,2,*] Jon Timothy M. Rivero[1,3,4] Valentin C. Dones III[1,3]**

[1]The Center for Health Research and Movement Science: A JBI Affiliated Group, University of Santo Tomas, Manila, Metro Manila, Philippines
[2]Department of Occupational Therapy, College of Rehabilitation Sciences, University of Santo Tomas, Manila, Metro Manila, Philippines
[3]Department of Physical Therapy, College of Rehabilitation Sciences, University of Santo Tomas, Manila, Metro Manila, Philippines
[4]The Ruth Foundation for Palliative and Hospice Care, Inc., Muntinlupa, Metro Manila, Philippines
*****Corresponding author: nrsantos@ust.edu.ph*

Rehabilitation professionals play essential but underrecognized roles in palliative care, navigating contrasts between restorative and palliative philosophies, while adapting their practice to prioritise patient comfort, participation, and dignity. Despite growing global recognition of rehabilitative palliative care, misconceptions, systemic barriers, and limited understanding of rehabilitation roles continue to restrict their involvement in palliative care delivery. Qualitative studies exploring the experiences and perspectives of rehabilitation professionals, including occupational therapists, physical therapists, and speech-language pathologists, working in palliative care settings were eligible. Studies using any qualitative design conducted in hospitals, hospices, home-based care, nursing homes, or community-based programs were included. Mixed-methods studies were considered when qualitative findings could be extracted. The review followed JBI methodology for systematic reviews of qualitative evidence. A comprehensive 3-step search strategy was conducted beginning February 20, 2025, across PubMed, CINAHL (EBSCOhost), Embase (Ovid), Web of Science, PEDro, and grey literature sources including ProQuest Dissertations & Theses, WHO resources, and Google Scholar. Search terms for palliative care were adapted from a validated search filter. Two reviewers independently screened titles, abstracts, and full texts, and appraised methodological quality using the JBI Critical Appraisal Checklist for Qualitative Research, with disagreements resolved through discussion or consultation with a third reviewer. Data were extracted using the JBI SUMARI tool and synthesized through JBI's meta-aggregation approach. Confidence in the synthesized findings was assessed using the ConQual approach based on dependability and credibility. The search and citation tracing identified 23,787 citations, of which 19,180 unique records were screened. One hundred eighty-seven reports were assessed, and studies from 16 countries met the inclusion criteria. A total of 277 findings were extracted, of which 275 were included in the synthesis, resulting in 42 categories and eight synthesized findings. These synthesized findings revealed that occupational therapists, physical therapists, and speech-language pathologists use flexible, integrative approaches that support autonomy, meaningful activity, and shared decision making with patients and families. Their interventions address physical, psychosocial, emotional, and spiritual dimensions of care, yet face systemic, cultural, and physician-centred barriers that limit their full integration into interdisciplinary teams. Professionals also reported emotional and relational demands as both burdens and sources of meaning, emphasizing the need to develop coping strategies and reflective practice. Advocacy, training, and policy reform must strengthen the recognition of the rehabilitation professionals' role, expand workforce competence, and embed rehabilitation within national palliative frameworks. Lastly, building compassionate communities through education and partnerships can normalise conversations about death and ensure sustainable, person-centred care beyond institutional settings. The overall ConQual scores for the synthesized findings were rated as low to moderate. Rehabilitation professionals contribute substantially to person-centred palliative care yet remain limited by structural, cultural, and organisational barriers. Strengthening their integration requires targeted education and training, advocacy, and policy reform that embed rehabilitation into national palliative care frameworks. The findings highlight the need to build workforce competence, enhance interdisciplinary collaboration, create standardized practice guidelines, and ensure timely and equitable referral pathways. Future research should further examine culturally responsive approaches to rehabilitative palliative care, assess the effectiveness of rehabilitation interventions across diverse settings, and identify strategies to improve workforce well-being and clarify professional roles.

**Keywords:** person-centred, interdisciplinary, hospice, community-based, occupation

# Design and Development of a Stress Detection System using an Intelligent Pen for the Student's Well-Being

**H.A. Kwan, J.P. Manalo, J.M. Santos, E. Torres, A.B. Valle, E. Roxas**

[1] *Department of Electronics Engineering, University of Santo Tomas, Manila, Philippines*
**Corresponding author: earoxas@ust.edu.ph*

Stress remains a prevalent mental health concern in modern society, necessitating innovative approaches for early detection and intervention. This study proposes the use of iSULAT (Intelligent Stroke Utilization, Learning, Assessment, and Testing), an intelligent pen system utilized to detect mental stress among engineering students at the University of Santo Tomas–Manila through handwriting analysis. The system integrates multiple instruments, including handwriting tasks, the Depression, Anxiety, and Stress Scale-21 (DASS-21), the Visual Analog Scale (VAS), and physiological data from a smartwatch, to provide a comprehensive assessment of stress. Handwriting features captured by the intelligent pen were correlated with physiological indicators, specifically Heart Rate Variability (HRV), and self-reported stress levels from the DASS-21 subscale, both of which correlated with the aspect of mental stress. A data collection of 150 participants was conducted in coordination with the UST Faculty of Engineering, scoping fourth-year engineering students across all the (6) engineering departments, wherein the entire dataset will be analyzed in its complete version and its aligned version. Each configuration was evaluated separately under the machine learning pipeline based on the three chosen algorithms: Support Vector Machine, Decision Tree, and Random Forest to assess how label consistency affects the model's performance. Throughout the entire procedure, it has been detected that air time was the most significant handwriting feature in predicting mental stress, particularly in tasks involving writing the words in capital letters. Among the evaluated models, the Random Forest (RF) algorithm demonstrated an accuracy percentage of 90.91% when identifying the best feature in detecting stress. Additionally, when checking the overall accuracy among all the proposed handwriting tasks, the Decision Tree (DT) algorithm obtained a balanced accuracy of 58.6%, supporting the feasibility of using an intelligent pen system for stress detection within the specified population.

# Perceptions of Generative AI Integration in Clinical Training Among Occupational Therapy Interns in Manila: A Qualitative Study

**Santos, Nikka Karla R.[1,2]*; Fariñas, Maria Ruby P.[1]; Behan, Sean James Eire A.[1]; Agraviador, Ryza Mikyla V.[1]; Buquing, Eladia Denise A.[1]; Diosanta, Josiah Myron A.[1]; Evangelio, Kristin Chloe L.[1]; Santiago, Ma. Dulce Regina D.[1]**

[1] Department of Occupational Therapy, University of Santo Tomas, Manila, Philippines
[2] Center for Health Research and Movement Science, University of Santo Tomas, Manila, Philippines
***** *Corresponding author: nrsantos@ust.edu.ph*

Generative Artificial Intelligence (GenAI) is increasingly being integrated in healthcare and rehabilitation education, with potential to support clinical decision-making, documentation, intervention planning, and student learning [1]. Its relevance is expanding in occupational therapy (OT), yet evidence on how GenAI is perceived and used in Philippine OT clinical training remains limited [2, 3]. This study examined the perceptions of OT interns in Manila, Philippines, regarding the integration of GenAI into clinical training, including perceived benefits, challenges, ethical considerations, and preferred approaches to integration, guided by the Diffusion of Innovations Theory. An exploratory qualitative design was employed. Twenty-eight (28) OT interns from two universities in Manila, with at least two months of clinical exposure and prior knowledge or experience with GenAI, participated in four online focus group discussions via Google Meet. Data were analyzed using thematic analysis. Four major themes emerged: (1) GenAI functions as a helpful support tool, (2) perceived limitations of GenAI integration, (3) potential risks of GenAI, and (4) the status of GenAI integration. Interns reported using GenAI to support evaluation and intervention planning, documentation, and academic learning, contributing to perceived improvements in confidence, efficiency, and clinical preparation. However, they emphasized that GenAI cannot replace OT clinical reasoning because of occasional inaccuracies, limited contextualization, and difficulty capturing holistic, client-centered OT practice. Institutional and contextual barriers, particularly variable guidelines and limited relevance to the Filipino practice context, further shaped its use. Ethical concerns included risks to patient data, confidentiality, safety, and over-reliance, which may affect clinical reasoning and professional development. Despite these concerns, interns viewed GenAI integration as unavoidable and recommended structured training, clearer policies, supervised practice opportunities, and ethical guidance. GenAI is perceived as a valuable yet conditional support tool in OT clinical training. Responsible integration requires alignment with professional standards, institutional safeguards, and continued emphasis on clinical reasoning, ethical practice, and contextually responsive OT education.

**Keywords:** generative artificial intelligence; occupational therapy; clinical training; interns; clinical education; Philippines

## References

[1] Kaelin, V. C., Valizadeh, M., Salgado, Z., Sim, J. G., Anaby, D., Boyd, A. D., et al. (2022). Capturing and operationalizing participation in pediatric re/habilitation research using artificial intelligence: A scoping review. *Frontiers in Rehabilitation Sciences, 3*. https://doi.org/10.3389/fresc.2022.855240

[2] Kaelin, V. C., Nilsson, I., & Lindgren, H. (2024). Occupational therapy in the space of artificial intelligence: Ethical considerations and human-centered efforts. *Scandinavian Journal of Occupational Therapy, 31*(1), Article 2421355. https://doi.org/10.1080/11038128.2024.2421355

[3] Lee, S. A., & Park, J. H. (2025). Artificial intelligence in occupational therapy documentation: Chatbot vs. occupational therapists. *DIGITAL HEALTH, 11*. https://doi.org/10.1177/20552076251386657

# Subacromial Region Ultrasound Image Enhancement and Segmentation for Acromiohumeral Distance Measurement

**Engr. Glenn C. Virrey, MSc, ECT, Gwen Stefani C. Loreno, Sean Darrel Y. Mariano, John Angelo B. Tabang, Jerald Tyrone V. Tadena, Lance Adriane B. Wallit, Prof. Consuelo G. Suarez, MD, PhD, FPARM, Jan-Tyrone C. Cabrera, MD, FPARM, Mary Monica B. Bueno, MD, PHD, Asst. Prof. Maria Belinda Cristina C. Fidel, MD, FPARM**

*Graduate School, University of Santo Tomas, España, Manila, Philippines*
*Faculty of Engineering, University of Santo Tomas, España, Manila, Philippines*
*College of Rehabilitation Sciences, University of Santo Tomas, España, Manila, Philippines*
**Corresponding author: gcvirrey@ust.edu.ph*

Shoulder Impingement Syndrome (SIS) is a prevalent musculoskeletal disorder characterized by the compression of soft tissues within the subacromial space. The Acromiohumeral Distance (AHD) serves as a critical diagnostic parameter for SIS. This study proposes an automated image processing framework to enhance ultrasound (US) clarity, segment the subacromial region, and measure AHD. A multi-stage enhancement pipeline was developed, with the “Enhance6” configuration utilizing adaptive histogram equalization, low boundary leveling, and 20 iterations of gamma correction emerging as the optimal protocol. Results demonstrate a 90% improvement in Signal-to-Noise Ratio (SNR), increasing from 3.3 to 6.6. Notably, the framework intentionally reduced the Contrast-to-Noise Ratio (CNR) to preserve the natural appearance of bone interfaces, a feature clinically validated to improve anatomical landmark identification. The automated measurement system achieved strong agreement with manual methods, reducing mean absolute error by approximately 60%, providing a practical, low-cost diagnostic support tool for resource-constrained environments.