Motion-Based Educational Games: Using Multi-Modal Data to Predict Player's Performance

Serena Lee-Cultura, Kshitij Sharma, Sofia Papavlasopoulou, Michail Giannakos
Institutt for dataeknologi og informatikk
Norwegian University of Science and Technology, Norway
serena.leecultura@ntnu.no, kshitij.sharma@ntnu.no, spapav@ntnu.no, michailg@ntnu.no

***Abstract*—Multi-Modal Data (MMD) can help educational games researchers understand the synergistic relationship between player's movement and their learning experiences, and consequently uncover insights that may lead to improved design of movement-based game technologies for learning. Predicting player performance fosters opportunities to cultivate heightened educational experiences and outcomes. However, predicting player's performance utilising player-generated MMD during their interactions with educational Motion-Based Touchless Games (MBTG) is challenging. To bridge this gap, we implemented an in-situ study where 26 users, age 11, played 2 maths MBTGs in a single 20-30 minute session. We collected player's MMD (i.e., gaze data from eye-tracking glasses, physiological data from wristbands, and skeleton data from Kinect) produced during game-play. To investigate the potential of MMD for predicting player's academic performance, we used machine learning techniques and MMD derived from player's game-play. This allowed us to identify the MMD features that drive rapid highly accurate predictions of players' academic performance in educational MBTGs. This might allow us to provide real-time proactive feedback to the player to support them through their educational gaming experience. Our analysis compared two data lengths corresponding to half and full duration of the player's question solving time. We showed that all combinations of extracted features associated with gaze, physiological, and skeleton data, predicted student performance more accurately than the majority baseline. Additionally, the most accurate prediction of player's performance derived from the combination of gaze and physiological data for *both* full and half data lengths. Our findings emphasise the significance of using MMD for real-time performance prediction in educational MBTG and offer implications for practice.**

***Index Terms*—motion-based games, educational games, multi-modal data, education, prediction, machine learning**

I. Introduction and Motivation

Motion-Based Touchless Games (MBTG) utilise sensing technologies to capture, map and interpret, player's body movements [5] as game input. In the context of education, MBTGs have recently emerged as a promising interactive didactic approach in pursuit of sharpening children's cognitive abilities [48]. Notable studies report that MBTGs are an interesting, effective and engaging way to support children during learning [20]. Educational MBTG play engages the physical body in a multi-modal learning experience [25], making these games attractive candidates for the application of Multi-Modal Data (MMD) collection; with potential consequences for the analytics and design aspects of games. Exploration into this space might enrich researchers insights of how body movement may infer player's understanding of the learning content.

Parallel to this, predicting a player's learning performance offers researchers and teachers opportunities to provide superlative pedagogical experiences due to the advantageous ability to deliver fine-grained, temporal and personalised learner feedback. Example include identifying when learners require additional assistance [17], [19] (i.e., real-time helpful hints and constructive evaluation) and altering the method of instruction to better accommodate the learner's progress [9]. These didactic strategies support the cultivation of effective, efficient learning trajectories by minimising unproductive practice and illustrate the importance of identifying adequate ways to effectively predict the learner's performance.

Predicting learning performance early on in an activity offers several benefits for the design and support of the learning experience [3], [17]. For example, Giannakos et al. [17] demonstrated the feasibility of using an Empatica E4 wristband data to infer early predictions of student's learning experiences during university class sessions. Andrade et al. [3] explored the potential of using MMD captured through hand gestures for personalisation and prediction. Collectively, the aforementioned works foretell potential benefits derived from obtaining a deep understanding of how the synergistic relationship between movement and learning experiences might drive performance prediction. However, limited research has attempted to interlace these ideas by investigating the feasibility of predicting learning performance using MMD collected during user's interactions with MBTG.

This study aims to address shortcomings in the literature by using MMD and machine learning to identify important MMD features that drive rapid, highly accurate predictions of players' academic performance in MBTGs. Specifically, our research question asks: *How accurately can we predict player's learning performance (through the correctness of answers) using educational MBTGs and MMD, and is early prediction feasible?*

To explore these issues, we conducted an empirical study with 26 children playing two MBTGs for mathematics (i.e., arithmetic and geometry). We collected MMD from three devices; gaze data from eye-tracking glasses, physiological data from Empatica E4 wristband, and skeleton data from Microsoft Kinect. Next, we generated a data sets for each combination of the data streams originating from the three

978-1-7281-4533-4/20/$31.00 ©2020 IEEE

different features. That is, (1) gaze, (2) physiological, (3) skeleton, (4) gaze and physiological, (5) gaze and skeleton, (6) physiological and skeleton, and lastly (7) gaze, physiological and skeleton. Finally, we applied machine learning to each data-set twice; as defined by full and half duration of students' answer time. In particular, we make the following contribution:

- We present insights from an in-situ study that collected MMD during player's physical engagement with two different educational MBTGs.
- We show that MMD collection facilitates early prediction of children's performance playing educational MBTGs.
- We discuss how our findings can be used to inform pedagogical research and practice.

II. Related Work

A. Educational Motion-Based Touch-less Games

Though evolution of motion sensing technologies has only recently enabled the development of MBTGs, the literature illustrates that researchers consider movement as a powerful pedagogical tool in children's education [21]. In these games, the child enters into a virtual world in which they use their body to interact naturally with educational material purposed to playfully develop their cognitive skills [48]. Numerous educational MBTG have been proposed and implemented, spanning a multitude of domains, such as maths [23], [28], [41], science [27], [44], language development [20], [48], vegetation succession [1] and special educational needs [5], [6], [8]. Encouraging results suggest that the positive impact of MBTGs on learning is multifaceted; with studies indicating strengthened self-confidence while learning [15], [26], increased engagement [20], and more fluid problem solving ability [14]. Additionally, the use of MBTGs in learning has also been shown to foster creativity and facilitate collaboration strategies through playful learning experiences [48].

Moreover, research in MBTGs has seen much traction in the context of mathematics, with research permeating numerous branches such as arithmetic [43], algebra [23], calculus [34], and geometry [41]. Notable studies suggest that MBTGs might have a positive impact on player's maths learning experience; particularly concerning reduced anxiety [22], enhanced problem understanding [4], [41], and increased academic performance [26], [39], [44]. Moreover, computational thinking skilled may be amplified via introducing students to programming languages that exploit MBT technologies, such as Scratch for Kinect (i.e., Kinect2Scratch) [2], [12].

Collectively, these contributions demonstrate that educational researchers and teachers are beginning to consider MBTG as a viable solution by which to augment the current instructional approach [21], particularly concerning maths. Despite the indicated advantages of introducing playful movement into educational spaces, little research has been conducted to uncover exactly how MMD generated during user's MBTG play might inform research on improving game design and development (e.g., to utilise early predictions, offer MMD-based analytics dashboards), as well as learning design practices (e.g., inform the instructor in real time).

The following section provides an overview of researcher's exploration of MMD to direct performance prediction in educational settings. As well, we touch on its application to movement based learning technologies (i.e., games).

B. Multi-Modal Data-Based Predictions in Education

MMD has been used to predict performance and engagement in educational contexts in previous research [3], [18], [42]. These contexts vary from games [18], [25], to assessment systems [40], to adaptive systems [32], to collaborative systems [24]. However, one common factor in these studies is the use of multiple data streams (e.g., gaze, facial expressions, Electroencephalography (EEG), heart rate, log data) to predict and explain learning performance [18], [40], behaviour [3], [42] or experience [37], [38].

For example, in the terms of learning performance prediction, researchers used EEG and behavioural data (e.g., reaction time from click streams, [7]) to predict students' recall, or gestures, postures, and body movements to predict students' performance in repeating, recalling and association tasks [24]. In a project-based learning case, Spikol et al. [42] used objects created by the students in their respective projects, combined with students' positions, hand gestures, facial expressions, audio, video and interaction patterns with the physical computing platform, to predict the quality and correctness of their solution. Liu et al., ([29] and [30]) used the log data, audio and video. In all studies, the prediction of performance achieved was highly accurate. Additionally, a combination of EEG, gaze data, facial expressions and physiological data was used to predict performance in an adaptive test [40] and game-play (i.e., Pacman game) [18] with a "high" level of accuracy (i.e., error rate was close to 6% in both the cases).

Moreover, when it comes to predicting learning behaviour or experiences, several researchers have pointed out the capability of MMD to be able to predict and explain the learning processes or students' trajectories. In particular, full body motion-based features/measurements have been found to explain and predict students' short-term memory and recall (e.g using kinematics and skeleton data [24], [25]). Furthermore, hand movement was used to predict the quality of students' projects [42], [46]. Physiological data such as, electrodermal activity (EDA), Galvanic Skin Conductance (GSC), skin temperature and accelerometer data, was used to predict students' engagement [37], [47] and perceived satisfaction with the learning task [17]. These studies report several results showing that MMD can be successfully used to used to explain students' behavioural trajectories while they engage in the learning task.

III. The Motion-Based Touchless Games

This section presents a detailed account of the educational MBTGs used in this study, Marvy Learns and Sea Formuli.

A. Marvy Learns: A Geometry Sorting Game

Marvy Learns was used to sharpen the player's geometry skills by mapping 2D flattened shapes (i.e., shape-nets) to their 3D representations. In this game, the player helped a creature, Marvy, organise a collection of shape-net cards, by moving each card into the box labelled with the corresponding 3D shape names. The player's body movement was mirrored by Marvy, so arrangement of shape-net cards occurred as the player moved their body in physical space, performed a mid-air hand gesture to select a card, and then bent down to place the shape-net card in the proper box. For example, Figure 1b shows 4 shape-nets to be sorted: an unfolded green triangular prism (upper left), an unfolded pink cube and an unfolded red cube (lower left and upper right, respectively), and an unfolded pink tetrahedron. The player must read the box labels (Triangular Prism, Tetrahedron, and Cube), then for each shape-net, visualise the objects that results from folding the shape-net into a 3D shape and move the card to the box with that name. In this way, players became familiar with new concepts by associating the displayed items with the defined words on the boxes. Each game consisted of a single matching question with 6 cards to be sorted. In addition, Marvy Learns fosters logical and inductive thinking through practice of arranging and classifying objects.

B. Sea Formuli: An Arithmetic Operations Game

Sea Formuli focused on developing a player's algebraic thinking by practising arithmetic problems involving whole numbers, decimals and fractions. To solve a problem, a player needed to calculate the missing number (i.e., operand) or operator in an equation relating 3 terms. Each question was represented by baskets sitting on the ocean floor (see Figure 2a), paired with three potential answers characterised by swimming jellyfish labelled with an operator or operand. The player was represented by a photo-realistic avatar that mimicked the complete range of their movements. To answer a question, a player selected a jellyfish by performing a mid-air hand gesture, and then moving their body by bending down to place the jellyfish into the empty basket to complete the equation. Figure 2a illustrates the decimal addition question: $4.02 + \_\_ = 8.12$, with potential answers: 4.1, 6.36, and 6.07. The player must perform the mental calculation, gesture to select the jellyfish labelled 4.1, and then move their body (and the selected jellyfish) into the empty basket on the ocean floor. Each game play session consisted of 5 multiple choice maths questions.

IV. Methodology

A. Context

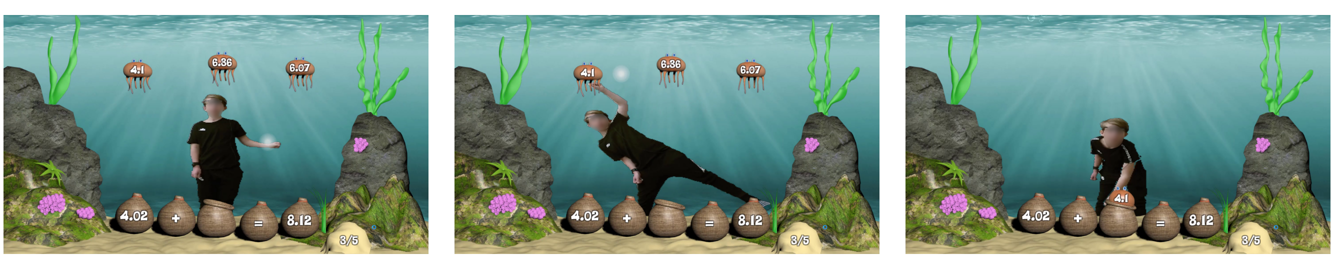
Our study was conducted in on a grade 6 class at Ila Skole in Trondheim, Norway in fall of 2019. Children volunteered to participate upon receiving a thorough explanation of the study by researchers and their maths teacher. The study was conducted by the researchers/authors, in a room dedicated strictly to the experiment. The room was arranged to accommodate two MBTG play session setups running in parallel.

B. Participants

Our sample was composed of 26 typically developing children (10 F, 16 M) with an average age of 10.95 years ($SD = 0.21$ years). All children participated in 6 game play sessions, for which they received a gift card for their participation. All procedures were approved by the Norwegian Centre for Research Data, (i.e., a national human research ethics organisation) and all children and their guardians provided verbal/written informed assent/consent, respectively, prior to participation.

(a) player is presented with geometric shapes to sort into labelled boxes using the avatar, Marvy. (b) player gestures and Marvy's body moves in sync to select the geometric shape. (c) player bends down to place the selected geometric shape into one of the labelled boxes.

Fig. 1: A player gesturing through a Marvy Learns game problem.



(a) player is presented with an arithmetic problem to solve using a photo-realistic avatar. (b) player gestures and the avatar mimics the player to select the jelly fish. (c) player bends down to guide the selected jellyfish to the empty basket completing the equation.

Fig. 2: A player gesturing through a Sea Formuli game problem.

C. Procedure

We conducted a study to investigate learner prediction through the application of MMD to children's engagement with two maths MBTGs. Upon receiving legal guardian consent and children's assent, children were given an Empatica E4 wristband and pair of Tobii eye-tracking glasses to wear. Children played 6 game play sessions; 3 consecutive sessions of Marvy Learns and 3 consecutive sessions of Sea Formuli. For each game, the 3 sessions consisted of a practice round, in which researchers assisted the children in understanding the associated game's objective and rules, and two non-practice sessions. In total, the 6 MBTG play sessions lasted between $20-30$ minutes. We ensured a balanced game order. None of the children had prior exposure to MBTGs.

D. Data Collection

We collected MMD from three different sources: gaze data from eye-tracking glasses, physiological data (with sensors for heart rate, blood-pressure, temperature and EDA levels) from wristbands, and skeleton data from Microsoft Kinect.

Eye-tracking: To capture children's gaze data, we used Tobii eye-tracking glasses at 50Hz sampling rate and one-point calibration. Tobii glass controller software recorded video documenting the child's field of view, via objective camera built into the nose-bridge of the glasses. Video resolution was 1920x1080 at 25 Frames Per Second (FPS).

Empatica E4 wristbands: Children's wristband data captured 4 different variables: Heart Rate Variability (HRV) (1Hz), EDA (64Hz), skin temperature (4Hz), and Blood Volume Pulse (BVP) (4Hz).

Kinect Skeleton: Kinect sensor was used to capture skeleton data, recorded at a a sampling rate of 1Hz. It consisted of the 3D position for 25 joints shown in the Figure 3.

E. Data Pre-processing

Eye-tracking: Fixations and saccades were identified using Tobii's default algorithm (for details please see [33]). A filter was applied to remove raw gaze points that classified as blinks. Pupil dilation is highly susceptible to personal and contextual biases. For example, time of day, screen-brightness, physical/medical health conditions, the child's age, gender, age, amount of sleep, and so on. To accommodate for this, we used the first 30 seconds of gaze data to normalise pupil dilation, effectively removing subjective and contextual biases. Further normalisation was obtained using the darkest (i.e., set to maximum) and brightest (i.e., set to minimum) screen shots obtained from the player's complete interaction, to account for screen brightness.

Empatica E4 Wristband: To remove unwanted spikes in the data, a simple smoothing function was applied to the time series of the four data streams obtained by the Empatica E4 wristband. Data was sectioned into windows, where a "window" refers to a time segment containing 100 successive data points. Our function examined consecutive windows in the time series and calculated a running average accordingly. Successive windows contained a 50 sample overlap. Like the

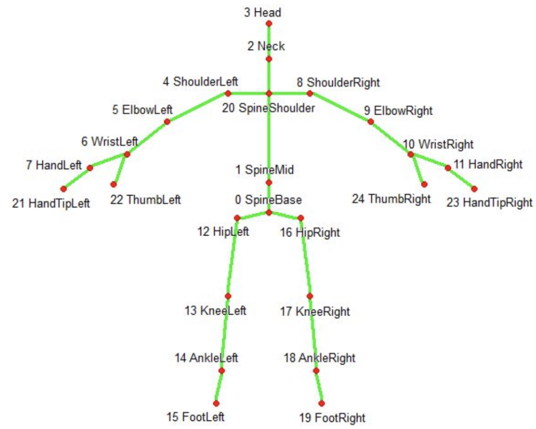


Fig. 3: The 25 joints captured by the Kinect skeleton data.

gaze data, the physiological data obtained by the wristband (namely, EDA, HRV, BVP and skin temperature), is highly susceptible to personal and contextual biases; such as time of day, pre-existing physical health conditions, the player's gender, age, amount of sleep, and so on. We normalised these features using the initial 30 seconds of the data streams to remove the subjective and contextual bias from the data.

Kinect Skeleton: No pre-processing was required.

F. Dependant variable: Correctness of answers

In our analysis, learning performance was computed based on the correctness of answer to a maths question during gameplay progression with the two MBTGs, (i.e., the dependant variable which also acted as the predicted variable). Correctness of answer is a binary variable (i.e., possible values or correct or incorrect). For the purpose of evaluating prediction quality, the "correct" class is the "positive" class.

G. Feature Extraction

We computed a total of 725 features from the different data streams: gaze, physiological (EDA, HRV, BVP, skin temperature), and 25 features from the Kinect joints. See Table I.

H. Feature Fusion

We computed seven combinations of features defined by the different data streams allocated to each input device: **gaze** (eye-tracking glasses), **physiological** (Empatica E4 wristband), **skeleton** (Kinect data), **gaze and physiological**, **gaze and skeleton**, **physiological and skeleton**, and **all** (gaze and physiological and skeleton). This allows us to explore the potential of each stream and their combinations.

I. Data Partition

We compared two data lengths: **full** and **half**. For the full data-set (mean duration = 94.75 seconds, SD = 61.20), we computed the features from the entire duration of one question (i.e., from the moment the question is displayed until an answer is logged). For the half data-set we used the initial 50% of the duration of one question (mean duration = 47.37 seconds, SD = 30.60).

TABLE I: Features Extracted

Wristband features (28 for four streams)	
Value histogram	Mean, SD, kurtosis and skewness of value histogram
Spectral histogram	Mean, SD, kurtosis and skewness of frequency histogram
ARMA	Auto-regressive moving average: maps the current value to the history of time series.
GARCH	Generalized Auto-regressive conditional heteroskedasticity: maps the current variance to the historical variance of time series and the heterogeneity of the appearance of the values
Joint distance travelled features (24 for 25 streams)	
Value histogram	same as above
ARMA	same as above
GARCH	same as above
Eye-tracking features (25)	
Pupil diameter	Mean, SD, kurtosis and skewness
Fixation duration	Mean, SD, kurtosis and skewness
Saccade	ratio of forward to backward saccades ratio of global to local saccades velocity (Mean, SD, kurtosis and skewness) amplitude (Mean, SD, kurtosis and skewness) duration (Mean, SD, kurtosis and skewness)
Events	Number of fixations, number of saccades, fixation to saccade ratio

J. Prediction using Ensemble Learning

To identify how each of the seven data combinations predict learning performance, we divided the complete data-set into training and testing subsets, retaining 15% data for testing (see details in the next section). Then, we used ensemble learning with model trees, support vector machines and Gaussian process modelling to predict the players' learning performance using the MMD. Ensemble models in machine learning combine the decisions from multiple models to improve the overall performance. In this paper, we combine predictions from seven different algorithms: Support Vector Machines [13] with linear, radial and polynomial kernels; Gaussian process models [45] with linear, radial and polynomial kernels; and M5 model trees. These methods are designed to improve the stability and accuracy of machine learning algorithms. One way of using the results from multiple models is to use a weighted average from all the prediction algorithms. The weights for individual prediction are considered based on their accuracy during the validation phase.

K. Training, validation, testing

First, we kept 15% samples (based on the participant IDs) aside for **out of sample testing**. Next, we performed training and validation on the remaining 85% samples. Validation was completed using the **leave one participant out** scheme (in each validation iteration samples corresponding to one participant were kept for validation). Finally, we tested the best model from the cross-validation phase on the out of sample test data set.

We observed our data-set to be heavily unbalanced. Particularly, it contains four times more correct answers than incorrect answers. To account for this we applied SMOTE (Synthesizing Minority Oversampling Technique) [31]. We implemented a SMOTE strategy by identifying the three nearest neighbours for each original point of the minority class and then adding three new (synthetic) points. The three new points were generated using the mean of the original point's three closest neighbours and then adding/subtracting 50% of standard deviation of the three neighbours to/from the mean.

L. Evaluation

We used the following metrics to evaluate the performance of the ensemble classifier: **Precision** = TP / (TP + FP); **Recall** = TP / (TP + FN); **Accuracy** = (TP + TN) / (TP + TN + FP + FN); **F1 score** = 2TP / (2TP + FP + FN). Where, TP = true positive; FP = false positive; TN = true negative; FN = false negative. For the purpose of evaluating the prediction quality the "correct" class is the "positive" class. For the baseline prediction, we selected the "majority class baseline", rather than the "random allocation baseline" due to the skewed nature of our data-set.

V. Results

Table II shows the results of predicting correctness of answers using the seven different combination of MMD features coming from gaze data, physiological data (EDA, HRV, BVP, temperature), and skeleton data, for both the full and half time segments. We observe that all seven MMD feature combinations perform better than the majority class baseline (F1 = 0.853, Figure 4), in both the full and half data length scenarios.

For the full data length case, the feature combinations of gaze and physiological data result in the highest F1-score (0.946, Figure 4) closely followed by the combination of features from all data sources (F1 = 0.9344, Figure 4). Whereas, the features extracted from the skeleton provided the worst F1-score (0.8917, Figure 4).

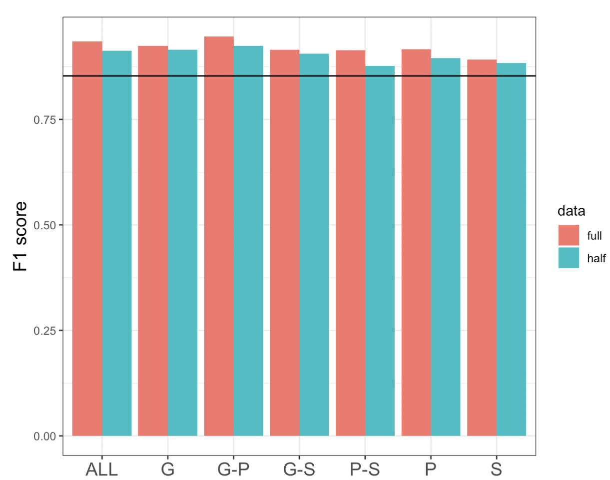


Fig. 4: F1-scores using different combinations of features from gaze (G), physiological (P; EDA, BVP, HRV, skin temperature) and skeleton (S) data. Two data lengths in time were compared: full and half.

TABLE II: Prediction results calculated using different combinations of MMD features from gaze (G), physiological (P; EDA, BVP, HRV, skin temperature) and skeleton (S) data. Two data lengths in time were compared: full and half.

Data	Features	Precision	Recall	Accuracy	F1-score
Full	**G**	0.9243	0.9243	0.8875	0.9243
	P	0.9090	0.9243	0.875	0.9166
	S	0.9196	0.8655	0.8437	0.8917
	G-P	0.9344	0.9579	0.9187	0.9460
	G-S	0.9304	0.8991	0.875	0.9145
	P-S	0.9380	0.8907	0.875	0.9137
	All	0.912	0.9579	0.9	0.9344
Half	**G**	0.9230	0.9075	0.875	0.9152
	P	0.9279	0.8655	0.85	0.8956
	S	0.9035	0.8655	0.8312	0.8841
	G-P	0.92 44	0.9243	0.8875	0.9244
	G-S	0.9292	0.8823	0.8625	0.9051
	P-S	0.9104	0.8403	0.8250	0.8772
	All	0.9016	0.9243	0.8688	0.9129
Majority baseline		**0.7438**	**1**	**0.7438**	**0.853**

Similarly in the half data length case, the combination feature of gaze and physiological data also result in the highest F1-score (0.9244, Figure 4), as well, again closely followed by the combination of features from all data sources (F1 = 0.9129, Figure 4). However when using only half of the data to calculate the prediction, the physiological and skeleton feature combination resulted in the worst F1-score (0.8772, Figure 4).

VI. Discussion

We observe that the combination of MMD features derived from gaze and physiological (EDA, BVP, HRV, temperature) data provide the most accurate prediction for both the full and half data length (see the F1-score column of Table II). A plausible explanation for this might be attributed to gaze's capacity to capture problem solving behaviour, such as cognitive load [10], attention [35] and anticipation [36]. Moreover, previous research [17] has shown that the four data streams included in the Empatica E4 wristband data have demonstrated the ability to capture constructs, such as, stress and physiological arousal. Such constructs (e.g., cognitive load, attention, stress) relate to problem solving behaviour. Thus, their capture aided us in achieving the reported prediction performance.

On the other hand, including the features provided by the skeleton data *does not* provide advantages. Rather, these data reduce prediction quality. This can be observed in Table II, by noting that in all occurrences, combining skeleton features with any other feature set always results in an reduced F1-score. For example, combining skeleton features with gaze features decreases the F1-score from 0.9243 (gaze) to 0.9145 (gaze and skeleton data) when using the full data length. Similarly, including the skeleton features with physiological features from the Empatica E4 wristband also degrades prediction accuracy by lessening the F1-score from 0.8956 (physiological data) to 0.8772 (physiological and skeleton data) when using the half data length. This indicates that skeleton data does not capture useful behaviour related to children's problem solving strategies. However, using the skeleton data to support the interaction (i.e., to play the game) during learning is important as MBTG have been found to a playful and engaging [20], [25].

One might conclude that in both the full and half data length scenarios, the difference between the most and least accurate predictions is negligible. By directly comparing the F1-score values corresponding to most and least accurate predictions, this may indeed seem convincing. In the full length data case, the least and most accurate prediction values are 0.8917 and 0.9460, respectively. Whereas, in the half data length case, least accurate prediction value is 0.8772 , and the most accurate is 0.9244. Hence, *directly* comparing these values, the resulting prediction accuracy improvements for the full and half data length scenarios are 0.0543 and 0.0472, respectively. However, such comparisons construct a false narrative, as the actual improvements to consider are calculated *relative* to the majority baseline (0.8530). Ergo, the least accurate prediction in the full and half data length cases are 0.0387 and 0.0242. Whereas, the most accurate predictions are 0.0930 and 0.0741. These represent improvements[1] of 240% and 306% in the full and half data scenarios, respectively.

One of the key findings of this contribution is utilising the capabilities of MMD to conduct early predictions during children's engagement with educational MBTGs. Table II indicates that using half (as oppose to full) data lengths to extract features and predict correctness, imposes relatively small decrements in the F1-scores (and all the other quality indicators). For example, for the gaze and physiological data feature combination, the half data length scenario improves the majority baseline prediction by 80% of that for the full data length scenario. To clarify, reducing the data set by half, only reduces the prediction quality by 20%. This illustrates the power of using partial data for early predictions, which leads to research implications regarding the ability to provide proactive and real-time feedback. Additionally, MMD provides clear benefit over alternative conventional methods used to capture children's learning experiences, such as self-reported data (e.g. questionnaires). In many cases, questionnaires can be inappropriate instruments for children with special needs or from different populations [17]. However, different MMD can be collected quickly, unobtrusively, universally and in a non-invasive manner while at simultaneously enabling high frequency user activity tracking.

VII. Implications

Wearables Support Academic Performance Prediction: A primary outcome to emerge from our analysis was the use of wearables as a driving force towards accurate prediction of children's maths performance resulting from MBTG play. Specifically, we identify the unification of data derived from the Empatica E4 wristband and eye-tracking glasses to determine the most accurate predictions in both the full and half time scenarios (see Table II, column F1-score). Notably, these wearables also greatly outperform the majority baseline on an

[1] $\text{improvement} = 100 * (\text{ best F1 } - \text{ baseline }) / (\text{ worst F1 } - \text{ baseline })$

individual basis, offering flexibility of use (i.e., to be used separately or together).

Advantages of Early Prediction: We observed that early prediction, using partial data length as measured by time, substantially outperformed the majority baseline prediction algorithm for all feature combinations (see Table II). Specifically, our analysis indicated that performance prediction with F1-score of 0.9244 (gaze and physiological data) could be conducted within the first 47.38 seconds (SD = 61.20 seconds) of children's engagement with the current question. Correspondingly, we identified two feedback design implications that enhance the real-time capability of artificial agent driven MBTG technologies. First, early prediction could be exploited to direct rapid proactive feedback (i.e., hints), thereby scaffolding children's learning as they develop, learn and master new skills during game play, as well as provide opportunities for reflection throughout the learning process. Second, early prediction provides the artificial agent additional optimisation time to prepare/select an appropriate follow-up question to present to the child, where appropriateness is determined by interplay of the child's learning goals and their academic understanding. Thus, the artificial agent might adjust the level of problem difficulty to keep the child in a state of flow (i.e., challenged, but not overwhelmed by questions posed). Lastly, the time saved by using early prediction (i.e., amount of time between prediction and logging of answer) translates to additional time for algorithm's optimisation purposes, which might yield better results.

Using Adequate Performance Predictors: Our results showed that use of body movement (i.e., skeleton data) to predict children's performance is not as accurate as the remaining extracted features (gaze and physiological data) included in our analysis. Therefore, grounded in our results and our specific context, we do not recommend that skeletal data be used for predicting for children's correctness or providing feedback. However, as previously mentioned, body movement's beneficial contribution to engage children during MBTG play is not diminished. This finding, also supports the fact that different data sources may have different purposes, possibly depending on the context of the educational activity (e.g., subject material). Encouraging the use of MMD will provide evidence for finding the appropriate context specific features for prediction.

Value for Educators in Classrooms: One implication revealed from our results is the benefit of using of MMD and early prediction within traditional classroom settings. Educators require technologies that will allow them to achieve effective educational practice. Thus, providing adequate information and prediction of children's performance on a given task, specially within 47.38 seconds ($SD = 61.20$ seconds), can result in identifying their learning and offer them appropriate guidance and personalised support. Moreover, the features used in our study can be combined or extended with additional MMD derived from different sources; for example, video recordings which can show children's affective state and contribute to explain more aspects of the learning experience, gaining a more holistic view.

Personalised Monitoring: Combining different data sources and predicting children's performance offers new possibilities for the development of technologies that support children in advancing their performance. More precisely, each (or even multiple) children can have a profile documenting their progress. This will allow them to track their advancement over time, see the potential changes in their own performance and act accordingly. Self-monitoring may possibly motivate them and advance their performance.

VIII. Conclusions & Future Work

In this paper, we motivate our research question with relevant literature and present an in-situ study that investigates the application of MMD collection to player's interactions with education MBTGs for the purpose of performance prediction. We considered both half and full question answering durations to determine whether the predicted outcomes justified a decrease in prediction accuracy. We conclude that the feature combination of gaze and physiological MMD provide the most accurate predictions. As well, we show the feasibility of early prediction. Though our learning context was limited to mathematics, our results echo and augment the findings of Giannakos et al. [16], [17], by showing that gaze and physiological from wristbands data can afford highly accurate early estimates of student performance, in the context of educational MBTGs. We note that the Kinect is no longer supported by Microsoft, however there exist other motion-based sensing technologies from which skeleton data can be extracted [11], and in no way are our results platform dependant. Lastly, we direct research and practice focused on MBTGs, by providing implications for learner experience design involving MMD in the context of education. Our findings emphasise the need for additional studies into investigate MMD within MBTGs, in efforts to create more fluid, reflective, and supportive educational experiences for both learner and instructor, particularly concerning learner feedback.

References

[1] T. Adachi, M. Goseki, K. Muratsu, H. Mizoguchi, M. Namatame, M. Sugimoto, F. Kusunoki, E. Yamaguchi, S. Inagaki, and Y. Takeda. Human sugoroku: full-body interaction system for students to learn vegetation succession. In *Proceedings of the 12th International Conference on Interaction Design and Children*, pages 364–367. ACM, 2013.

[2] I. Altanis and S. Retalis. A multifaceted students' performance assessment framework for motion-based game-making projects with scratch. *Educational Media International*.

[3] A. Andrade, J. A. Danish, and A. V. Maltese. A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3):18–46, 2017.

[4] R. Angotti and I. Bayo. Making kinections: Using video game technology to teach math. In *Prato CIRN Community Informatics Conference*, volume 2012, 2012.

[5] L. Bartoli, C. Corradi, F. Garzotto, and M. Valoriani. Exploring motion-based touchless games for autistic children's learning. In *Proceedings of the 12th international conference on interaction design and children*, pages 102–111. ACM, 2013.

[6] L. Bartoli, F. Garzotto, M. Gelsomini, L. Oliveto, and M. Valoriani. Designing and evaluating touchless playful interaction for asd children. In *Proceedings of the 2014 conference on Interaction design and children*, pages 17–26. ACM, 2014.

[7] M. Beardsley, D. Hernández-Leo, and R. Ramirez-Melendez. Seeking reproducibility: Assessing a multimodal study of the testing effect. *Journal of Computer Assisted Learning*, 34(4):378–386, 2018.
[8] A. Bhattacharya, M. Gelsomini, P. Pérez-Fuster, G. D. Abowd, and A. Rozga. Designing motion-based activities to engage students with autism in classroom settings. In *Proceedings of the 14th International Conference on Interaction Design and Children*, pages 69–78. ACM, 2015.
[9] P. Blikstein and M. Worsley. Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2):220–238, 2016.
[10] R. Buettner. Cognitive workload of humans using artificial intelligence systems: towards objective measurement applying eye-tracking technology. In *Annual Conference on Artificial Intelligence*, pages 37–48. Springer, 2013.
[11] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.
[12] C.-K. Chang and Y.-T. Tsai. Pair-programming curriculum development of motion-based game for enhancing computational thinking skills. In *2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI)*, pages 284–287. IEEE, 2018.
[13] O. Chapelle and V. Vapnik. Model selection for support vector machines. In *Advances in neural information processing systems*, pages 230–236, 2000.
[14] M. I. Contreras, C. G. Bauza, and G. Santos. Videogame-based tool for learning in the motor, cognitive and socio-emotional domains for children with intellectual disability. *Entertainment Computing*, 30:100301, 2019.
[15] J. Edwards, S. Jeffrey, T. May, N. J. Rinehart, and L. M. Barnett. Does playing a sports active video game improve object control skills of children with autism spectrum disorder? *Journal of sport and health science*, 6(1):17–24, 2017.
[16] M. N. Giannakos, S. Papavlasopoulou, and K. Sharma. Monitoring children's learning through wearable eye-tracking: The case of a making-based coding activity. *IEEE Pervasive Computing*, pages 1–12, 2020.
[17] M. N. Giannakos, K. Sharma, S. Papavlasopoulou, I. O. Pappas, and V. Kostakos. Fitbit for learning: Towards capturing the learning experience using wearable sensing. *International Journal of Human-Computer Studies*, 136:102384, 2020.
[18] M. N. Giannakos, K. Sharma, I. O. Pappas, V. Kostakos, and E. Velloso. Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, 48:108–119, 2019.
[19] A. Hellas, P. Ihantola, A. Petersen, V. V. Ajanovski, M. Gutica, T. Hynninen, A. Knutas, J. Leinonen, C. Messom, and S. N. Liao. Predicting academic performance: a systematic literature review. In *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, pages 175–199, 2018.
[20] B. D. Homer, C. K. Kinzer, J. L. Plass, S. M. Letourneau, D. Hoffman, M. Bromley, E. O. Hayward, S. Turkay, and Y. Kornak. Moved to learn: The effects of interactivity in a kinect-based literacy game for beginning readers. *Computers & Education*, 74:37–49, 2014.
[21] H.-m. J. Hsu. The potential of kinect in education. *International Journal of Information and Education Technology*, 1(5):365, 2011.
[22] K. Isbister, M. Karlesky, J. Frye, and R. Rao. Scoop!: a movement-based math game designed to reduce math anxiety. In *CHI'12 extended abstracts on human factors in computing systems*, pages 1075–1078. ACM, 2012.
[23] K. Johnson, J. Pavleas, and J. Chang. Kinecting to mathematics through embodied interactions. *Computer*, 46(10):101–104, 2013.
[24] M. J. Junokas, R. Lindgren, J. Kang, and J. W. Morphew. Enhancing multimodal learning through personalized gesture recognition. *Journal of Computer Assisted Learning*, 34(4):350–357, 2018.
[25] P. Kosmas, A. Ioannou, and S. Retalis. Moving bodies to moving minds: a study of the use of motion-based games in special education. *TechTrends*, 62(6):594–601, 2018.
[26] M. Kourakli, I. Altanis, S. Retalis, M. Boloudakis, D. Zbainos, and K. Antonopoulou. Towards the improvement of the cognitive, motoric and academic skills of students with special educational needs using kinect learning games. *International Journal of Child-Computer Interaction*, 11:28–39, 2017.
[27] C. Kynigos, Z. Smyrnaiou, and M. Roussou. Exploring rules and underlying concepts while engaged with collaborative full-body games. In *Proceedings of the 9th International Conference on Interaction Design and Children*, pages 222–225. ACM, 2010.
[28] E. Lee, X. Liu, and X. Zhang. Xdigit: An arithmetic kinect game to enhance math learning experiences. *Retrieved February*, 14:2013, 2012.
[29] M. Liu, E. McKelroy, S. B. Corliss, and J. Carrigan. Investigating the effect of an adaptive learning intervention on students' learning. *Educational technology research and development*, 65(6):1605–1625, 2017.
[30] R. Liu, J. Stamper, J. Davenport, S. Crossley, D. McNamara, K. Nzinga, and B. Sherin. Learning linkages: Integrating data streams of multiple modalities and timescales. *Journal of Computer Assisted Learning*, 35(1):99–109, 2019.
[31] L. Lusa et al. Smote for high-dimensional class-imbalanced data. *BMC bioinformatics*, 14(1):106, 2013.
[32] P. Mock, P. Gerjets, M. Tibus, U. Trautwein, K. Möller, and W. Rosenstiel. Using touchscreen interaction data to predict cognitive workload. In *Proceedings of the 18th ACM international conference on multimodal interaction*, pages 349–356, 2016.
[33] A. Olsen. The tobii i-vt fixation filter: Algorithm description [white paper]. *Retrieved from Tobii Technology from http://www.tobiipro. com/siteassets/tobiipro/learn-and-support/analyze/how-do-we-classify-eyemovements/tobii-pro-i-vtfixation-filter. pdf*, 2012.
[34] H. M. O. Orona, G. S. Maldonado, and N. P. S. Martínez. Kinect team: Kinesthetic learning applied to mathematics using kinect. *Procedia Computer Science*, 75:169–172, 2015.
[35] S. Papavlasopoulou, K. Sharma, M. Giannakos, and L. Jaccheri. Using eye-tracking to unveil differences between kids and teens in coding activities. In *Proceedings of the 2017 Conference on Interaction Design and Children*, pages 171–181. ACM, 2017.
[36] I. Pappas, K. Sharma, P. Mikalef, and M. Giannakos. Visual aesthetics of e-commerce websites: An eye-tracking approach. 2018.
[37] H. J. Pijeira-Díaz, H. Drachsler, P. A. Kirschner, and S. Järvelä. Profiling sympathetic arousal in a physics course: How active are students? *Journal of Computer Assisted Learning*, 34(4):397–408, 2018.
[38] M. Raca and P. Dillenbourg. Classroom social signal analysis. *Journal of Learning Analytics*, 1(3):176–178, 2014.
[39] S. Retalis, T. Korpa, C. Skaloumpakas, M. Boloudakis, M. Kourakli, I. Altanis, F. Siameri, P. Papadopoulou, F. Lytra, and P. Pervanidou. Empowering children with adhd learning disabilities with the kinems kinect learning games. In *European Conference on Games Based Learning*, volume 2, page 469. Academic Conferences International Limited, 2014.
[40] K. Sharma, Z. Papamitsiou, and M. Giannakos. Building pipelines for educational data using ai and multimodal analytics: A "grey-box" approach. *British Journal of Educational Technology*, 2019.
[41] C. P. Smith, B. King, and J. Hoyte. Learning angles through movement: Critical actions for developing understanding in an embodied activity. *The Journal of Mathematical Behavior*, 36:95–108, 2014.
[42] D. Spikol, E. Ruffaldi, G. Dabisias, and M. Cukurova. Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning*, 34(4):366–377, 2018.
[43] V. Thakkar, A. Shah, M. Thakkar, A. Joshi, and N. Mendjoge. Learning math using gesture. In *International Conference on Education and e-Learning Innovations*, pages 1–3. IEEE, 2012.
[44] C.-H. Tsai, Y.-H. Kuo, K.-C. Chu, and J.-C. Yen. Development and evaluation of game-based learning system using the microsoft kinect sensor. *International Journal of Distributed Sensor Networks*, 11(7):498560, 2015.
[45] C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.
[46] M. Worsley and P. Blikstein. Deciphering the practices and affordances of different reasoning strategies through multimodal learning analytics. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, pages 21–27, 2014.
[47] M. Worsley and P. Blikstein. Leveraging multimodal learning analytics to differentiate student learning strategies. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, pages 360–367. ACM, 2015.
[48] K. Yap, C. Zheng, A. Tay, C.-C. Yen, and E. Y.-L. Do. Word out!: learning the alphabet through full body interactions. In *Proceedings of the 6th Augmented Human International Conference*, pages 101–108. ACM, 2015.