An Innovative Model for Adaptive Learning Utilizing Biofeedback and Item Response Theory

László Gazdi[1*], Krisztián Pomázi[1], Máté Szabó[1], Bertalan Forstner[1]

[1] Department of Automation and Applied Informatics, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, H-1117 Budapest, Magyar Tudósok krt. 2., Hungary
* Corresponding author, e-mail: laszlo.gazdi@aut.bme.hu

Abstract

Measuring and providing efficiency of educational applications is a serious, open problem, which impacts the future of this expanding industry greatly. Reaching player engagement is a complex challenge, as it also depends on the given task and the mental state of the player. Researches answer this by using adaptive educational games. To reach the goal, however, knowledge about more parameters is required about the game tasks, the abilities of the player, his actual physiological state and performance as well. In this paper we present our results, which use a biofeedback based adaptive algorithm, and based on this, an innovative psychometric model to take a step towards maximizing user engagement.

Keywords

biofeedback, cognitive abilities, item response theory, multiple intelligence, classification

1 Introduction

People like doing something if they are enjoying it. Many activities, especially playing games can reach a point when the subjects are engaged and want to perform the activity continuously. Nowadays, in video games, things like visual effects, the reward system or the instant feedback over the performance can achieve this engagement easily. Players can learn moves or methods during the gameplay. This could make the games ideal candidates for educational purposes, if we can achieve the same easy learning effect. These games, that have educational purposes beside the entertainment, are called serious games [1]. Motivating the users during serious games is a key problem [2]. If the motivation is high, and the user enjoys the game, it can lead to longer and more effective playing sessions. For the most suitable gameplay, beside the ergonomic aspects, the appropriate level of difficulty is extremely important. It affects the mental state of the user, which, in turn, affects the performance. The theory of Flow [3] describes this approach. Typically, at least three relevant mental states are identified, based on Csíkszentmihályi's theory: boredom, frustration and engagement. The user is in the state of boredom if the level of the task is lower than the skill of the user. Frustration is defined as a state, where the task demands are higher than the skill of the user. Engagement is the ideal state, where the level of the task and the skill of

the user meet. This state leads to the Flow experience, where the efficiency is maximized. Many studies aim to introduce adaptive difficulty management in serious games [4]. The adaptiveness of these systems are based on the performance of the user in most cases, but for the ideal learning efficiency, the mental state of the user should also be considered.

To create an adaptive experience, adjusting the game difficulty based on the abilities and mental state of the player is found to be of key importance. In many games, gameplay can be divided into several well-defined tasks (steps or milestones). Each of these tasks has a difficulty level from which the difficulty of gameplay is determined. By giving the appropriate tasks for each player, their performance can be maximized, which enables more efficient learning and provides motivation.

In case the abilities of players can be measured, mathematical models can be used to estimate the probability that a player solves a given game task correctly. *Item Response Theory* (IRT) is particularly useful for this purpose [5-9]. In an educational game, each game task corresponds to a test item and each player corresponds to a subject in an IRT model.

In some cases, no estimate is available for the difficulty of game tasks. Fortunately, IRT models, being mathematical,

can be estimated in a lot of ways. Using model estimation, we can approximately measure the parameters of IRT items, which means, the difficulty of game tasks can be estimated using data from the responses of players.

By our theory, item parameter estimation can be simplified if we determine the ability of players via alternate means. One such method is by measuring a person's intelligence, based on the Multiple Intelligence Theory of H. Gardner. In his theory, Gardner states that intelligence can be differentiated into many inter-correlated modalities, rather than one single overlaying construct [19]. These modalities are defined as 9 individually describable and measurable intelligence fields. In this paper, we present an automated process to measure a number of the areas. We use the result of these automated measurement as the input of the IRT calculations.

We also provide a simplified method for estimating the difficulty of game tasks when the ability of players is treated as known. Our approach uses a modified version of IRT model estimation. Furthermore, we also propose an extension of simple IRT models, which can take the mental effort of subjects into account when determining the probability of solving a specific task by a specific user. Based on that information, we can manipulate the gameplay in order to keep the player in the state of Flow.

The rest of this paper provides a deeper insight into our system. The next section describes the related work and the theories, which our solution is based on. In the following chapter, the classification-based mental effort measurements are presented. After this, we show how cognitive games can be used to measure the ability of someone. The following chapters of this section describe the IRT model which uses mental effort and ability calculated by these cognitive games as parameters. A summary of our results and future works conclude our paper.

2 Related work
2.1 Measuring mental effort
The mental effort or cognitive workload of the user could be concluded from physiological parameters [10]. In our research, we measure mental effort through two physiological value: the hearth rate variability and the EEG data.

2.1.1 Heart rate
Heart rate is influenced by the activity of both the sympathetic and parasympathetic branch of the autonomic nervous system. An increase in sympathetic activity raises the hearth rate, while an increase in parasympathetic activity

lowers it. High mental effort levels cause increased heart rate, and decreased heart rate variability. Both time- and frequency domain-based analysis can be used to infer the mental effort of the user [11]. A parametric auto regression (AR) based power spectral density estimation and a moving average (MA) based time domain approach are primary used in the analysis.

We use the Low frequency (0.07-0.15 Hz) component as a measure of the mental effort [12]. Compared to a baseline value, a lower ratio would mean parasympathetic, while a higher would signal sympathetic dominance [13]. The aim of the MA method is to catch any changes in frequency (RR intervals).

2.1.2 Electroencephalography
EEG measures brain activity based on electric changes elicited by neurons. These, 10-100 μV signals can be measured with electrodes on the scalp [13]. In the last years, EEG became very common in human-computer interaction [15, 16]. There are many metrics to measure the mental workload based on EEG. Most studies use the power estimates of certain frequency bands (alpha, beta, delta or theta) [17]. Higher levels of load are accompanied by lower overall alpha power over activated cortical areas [18].

2.2 Item Response Theory
2.2.1 Basic concepts
Item Response Theory (IRT) [5-9] is a method for designing and scoring tests that consist of simple questions, which are called test *items*. The term item is generic: it can be used for a variety of question types, including multiple choice and open-ended questions. However, for an item, it should be easily determinable whether a given subject answered correctly.

IRT can be used to measure the *ability* of subjects taking the test. Ability is a latent trait, which means it is not a directly measurable characteristic of subjects. According to IRT, the test performance of subjects can be predicted or explained based on their ability level. Thus, ability is defined as a parameter of subjects that correlates with their performance on the test. Generally speaking, higher ability means higher probability that the subject gives a correct answer to the test item.

IRT defines mathematical models which are called *item response models*. *Unidimensional* IRT models, which are the most commonly used ones in practice, consider a single trait for subjects, the aforementioned ability. Another way to categorize IRT models is whether their

test items each have a single correct answer or not. In the first case, the items are called *dichotomous*. In the latter case, responses each have (a different) score value, meaning the items are *polytomous* [9, 19].

The common in all IRT models is that they define a function for each test item that is called the *item response function* (IRF). In unidimensional models, subject ability is the single input variable of this function. Most commonly, the IRF is a modified logistic function. In this case, the model is called a *logistic model*. An alternative formulation uses the cumulative distribution function (CDF) of the normal distribution. These models are sometimes called *normal ogive model*s. Nevertheless, in each case, the plot of the IRF (the *item characteristic curve*, ICC) is of a sigmoid shape.

IRT models differ in that how many factors they consider when characterizing test items. [20] The *one parameter logistic model* (1PL, sometimes also called the *Rasch model* [21]) defines a single parameter for items: the item *difficulty*, which determines at which ability value the midpoint of the ICC (the mean of the value for the lower and the higher asymptotes) is located.

The *two parameter logistic model* (2PL) adds *discrimination* as an item parameter, which characterizes the degree the item discriminates between subjects with different ability levels. Mathematically, discrimination is the slope for the tangent of the curve at its midpoint. 1PL assumes this degree is equivalent for all test items. Moreover, 1PL and 2PL do not consider that a subject can determine the correct answer for an item purely by guessing: in these models, guessing is part of the ability. The *three parameter logistic model* (3PL) takes *guessing* into account as the lower asymptote of the ICC.

For the 3PL, the probability of the correct response for a given dichotomous item can be given with the following formula: [9]

$$p\left(\theta\right) = c + \frac{1-c}{1+e^{-a(\theta-b)}}. \tag{1}$$

Here, θ denotes the ability of the subject solving the item, a denotes discrimination, b is the difficulty of the item [21], and c is the guessing parameter [23].

Additionally, a fourth parameter can be defined if one considers that even subjects with very high ability will have a probability of not answering an item correctly. In this case, the ICC of each item also has a variable upper asymptote. This version is rarely used in practice today, but it plays an important role in our research, as we will describe it later.

2.2.2 Estimating parameters of IRT models

When considering the traditional applications of IRT models in practice, the item and subject parameters are usually unknown at some stage of the model specification. When specifying the model, the item parameters need to be estimated. Typically, a random calibration sample from a target population is used for this purpose, where item and subject parameters are estimated simultaneously [5-7].

This estimation can be done with a variety of methods. In each case, the item and subject parameters are unobservable, thus a problem of indeterminacy arises. This is formally called the *identification problem*. Simply speaking, this means that certain transformations leave the IRF invariant when estimating two and three parameter models. In these cases, either the subject abilities or the item difficulties are fixed so that their mean is 0 and their standard deviation is 1.

One of the methods used for simultaneous parameter estimation is called *joint maximum likelihood estimation* (JMLE). Using JMLE, the values of item parameters and subject parameters need to be estimated in a way that they jointly maximize the value of the logarithm of the likelihood function. Variants of the iterative *Newton-Raphson procedure* or an alternative procedure, called *Fisher's method of scoring* can be used when performing this kind of estimation.

Other methods include *conditional maximum likelihood estimation* (CMLE) that can be used when sufficient statistics are available for the subject parameters which is only true in case of the one parameter model. *Marginal maximum likelihood estimation* (MMLE) [23] of the item parameters is carried out by integrating or summing over with respect to the subject parameters, and the function to be maximized is named the *marginal likelihood function* in this case. Two and three parameter models can also be estimated using MMLE. Finally, Bayesian methods can also be used for estimation.

After the item parameters are estimated from the calibration sample, they are treated as known in subsequent applications and item banks are constructed. The task is then to estimate the ability of subjects taking a test made from the items of the item bank. Simple *maximum likelihood estimation* (MLE) is most commonly used for this purpose. [5-7] Maximum likelihood estimators have some particularly useful properties like asymptotic normality and consistency. The standard error for the ability estimate can also be obtained using MLE which enables approximate confidence intervals to be calculated. In case of a

perfect or zero score, MLE fails. Bayesian methods can be considered as an alternative solution when this problem arises. Bayes estimators have smaller standard errors but require the specification of a prior belief regarding the ability of subjects.

2.3 Multiple Intelligence Theory

In his multiple intelligence theory, Gardner states that intelligence can be differentiated into many inter-correlated modalities, rather than one single overlaying construct [19]. He specifies nine fields of intelligence: musical-rhythmic, visual-spatial, verbal-linguistic, logical-mathematical, bodily-kinesthetic, interpersonal, intrapersonal, naturalistic and existential [26].

There are existing methods to measure cognitive abilities. One example is the Map of Interest Method, created by Eva Gyarmathy [27]. Based on the same Multiple Intelligence theory of Gardner, it uses questionnaires and self-evaluation to measure abilities. The subjects have to classify how much they agree with a given statement on a scale of 1 to 5, where the statements represent one of the 9 intelligence fields. The results can be used to determine the strengths and weaknesses of the individual where talent may manifest itself [28].

There are several other measurement methods based on the Multiple Intelligence Theory.

Profiling questionnaires are provided by other researches to measure multiple intelligence. In their research, Tirri, Nokelainen and Komulainen argue that sensitivities is a better word to call what Gardner calls intelligences [29]. Their analysis concludes that "there is no definite answer to the basic question, whether the multiple intelligences model can be confirmed in self-evaluated intelligence".

When comparing a learning strategy based on Gardners MI theory to a classic learning strategy, Mazaheri and Fatemis measurement shows that higher results can be achieved, if we take the students' preferred intelligence fields in account [30]. They conclude, "that e-learning strategy based on multiple intelligences helps students gain a basic understanding of scientific concepts since has been offered the curriculum in a meaningful and personalized manner for students were not adaptable in compared with traditional learning methods. MI has also been integrated to the learning system, with promising results" [31]. According to Hafidi and Lamia's research [32], "Experimental results show that the proposed system can precisely provide personalized activity recommendations

on-line based on learner abilities and responses, and moreover can accelerate learner learning efficiency and effectiveness." Learning Styles is another theory, which can be applied to enhance personalized learning. Denig's research suggests, that if we also want to impact the way which the students learn, applying Learning Styles can lead to better learning results [33].

As compared to these solutions, this paper presents a more universal and more automated approach by using a framework to design games. This creates the possibility to self-standardize the results and create a flexible and adaptive set of data.

Our solution uses the AdaptEd Framework [34-38]. The framework is based on a loopback model, which uses the live physiological signals and calculated values to finetune the running gameplay and optimize learning efficiency and performance. The framework takes care of data collection from the sensors, and it also processes the results, to provide useful and analyzable data.

It has four main components, as presented on Fig. 1:

1. The framework and the games run on the device of the player. These are individual applications, they can run separately. The framework runs in the background. It collects data, sent by the games, and sends messages to control the games.
2. The supervisor application runs on the device of the teacher. It is used not just to monitor, but to manipulate the gameplay. The teacher is able to start, pause, restart, mute, change the level or topic of the gameplay remotely. Moreover, she is able to modify the training set of the running algorithm.
3. The sensors provide the biofeedback data to the framework. They are typically wireless devices, connected by Bluetooth.
4. The server side stores the biofeedback and gameplay data, uploaded by the framework. Multiple visualizations are available about the data. The stored historical data can help in observing the development of a student.

3 Classification in mental effort measurement

We compare different classifiers on the measured heart rate and EEG data. The role of the algorithms is to identify the mental effort of the user. Based on this information, we would be able to tell, based on the IRT model, that the resolved task was easy or hard for the user, and can make recommendation to choose a harder or easier task from the item bank next. [37]
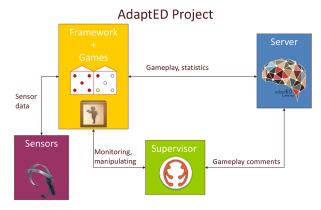
Fig. 1 System Architecture. — There are 4 main roles, which the architecture is divided into: the supervisor, the game, the framework and the server.

Fig. 2 Picture of an N-back gameplay: the red arrow shows the working classification (SVM) algorithm

3.1 Procedure

8 adult subjects participated in our study at the Department of Ergonomics and Psychology in Budapest University of Technology and Economics. All participants reported healthy. None of them had any diagnosed cardiovascular disorders. They were asked to not take any stimulants (energy drinks, coffee, cigarettes) or alcohol 8 hours before the tests.

Each recording began with a 2-minute-long relaxation phase. This data was the baseline for every physiological channel. After this phase, the participants got tetris, arithmetical and n-back tasks with multiple, predefined difficulty levels. We used these type of tasks, because it is easy to specify their difficulty with simple parameter changes. A session on one difficulty level took 2 minutes. Between the sessions, there were 1 minute long relaxation phases. During the tetris tasks, the participants had to play tetris with three different difficulty levels. The difference between the levels was the speed of the elements.

The N-back tests are widely used in cognitive neuroscience, for testing mental workload [16]. In this task, participants are presented with a sequence of letters, one at a time. One character was shown per second. They are asked to compare the current letter to one presented n items prior in the sequence. An exception is the 0-back test, where the user has to focus only on one letter presented in the beginning of the test. This is the target, and the stimulus is when it is matched with any letters of the sequence. During the 3-back, the current stimulus is a target when it matches the letter presented three letters ago. During the n-back test, the participants got 0-back, 1-back, 2-back and 3-back tasks.
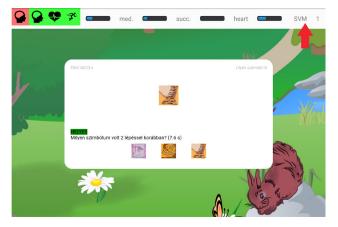
3.2 Data preparation

ECG (heart rate variability) was recorded with an HxM BT heart rate monitor. EEG data was recorded by an Emotiv Epoc EEG headset. The measurements were made with one second accuracy, but started at different time. To be able to use the measured physiological data, we needed to synchronize and transform them. We cleared the first few records from the data, until the start of the first relaxation phase. Every data was normalized with the baseline data, measured under the relaxation phase. This way, we could compensate the great personal differences, and got a normalized dataset. First, we made linear interpolation on the RR ECG data to be able to sample it once per 1000 ms. We used Burg's algorithm and SDNN algorithm to create the necessary dimensions for the classification. The prepared data on the ECG dimensions are presented on Fig. 3. The axis X shows the normalized data after used Burg's algorithm on the raw data, the axis Y shows
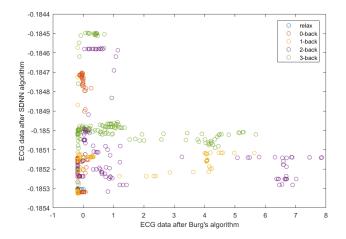

Fig. 3 Visualization of the data.

the normalized data after used SDNN algorithm. The different colors represent the different mental states. We can see, that the different mental states during the tasks can be separated well.

3.3 Algorithms

We used 10-fold cross validation for evaluating the data. The data set is separated to ten equal parts randomly. We used nine to train and one to test the algorithms. Every part and record is used nine times for training and once for testing. With this method, we could avoid any dispro-portion, that can be caused by the random choosing of training and test set.

We evaluated each training with a confusion matrix, and summarized the results. From the matrix, we were able to see all of the correct and false classifications. The whole procedure was repeated ten times from the begin-ning, the random separation of the data. This is necessary, in order to completely avoid any kind of influence of the random selection. So we used the 10-fold cross valida-tion ten times on the data set, but always with a differ-ent indexing. (We marked the nth indexing of the data set with IDS_n.) We take the average of the result in order to avoid the extreme high or low rates, that can be caused by a single selection.

We tested both classification algorithms capable and non-capable of multi-class classification. The multi-class classification can be created from binary-class classifica-tion, with the one-versus-one method. The algorithms clas-sify the vectors, and they vote for the label. Every vector gets the label with the most vote. In our measurements, we tested classification algorithms for binary use, however, some of them are capable for multiclass classification. We chose the 0-back and 3-back state for our data to classify.

3.3.1 Support Vector Machine

A support vector machine constructs a hyperplane in the space, which is used to separate the two classes, with the biggest margin. We tested a linear, a quadratic, a poly-nomial and a Gaussian Radial Basis Function kernel to decide, which works best for our dataset.

3.3.2 K-Nearest Neighbor

The K-Nearest Neighbor algorithm first measures the dis-tances of the k nearest vector from the new vector, and based on a rule, decides the class of the actual point. We used the Eucledian distance as the distance metric. As the assignment rule we tested the nearest and the consensus rule.

3.3.3 Logistic regression

The logistic regression is used to model the relationship between the dependent variable and the independent vari-able. In our case, the class of the vector was the dependent value. The EEG data and heart rate variability were the independent values.

3.4 Results

We evaluated the data 10 times, separately on IDS_1, IDS_2, … IDS_{10}. The differences were small between the data sets, so we present the average of these, as the overall results.

The confusion matrix of the summarized results with lin-ear support vector machine is shown at Table 1. We can see, that 1160 times from 1200 (96.67 %) the 1-back, and 1340 times from 1360 (98.53 %) the 3-back classification was cor-rect. The 1-back recognition was false at 0.017 % (20/1160). The false recognition rate at 3back is 0.029 % (40/1380).

We also tested the support vector machine with qua-dratic, polynomial and radial basis function kernel. The results are presented at Table 2. Each two rows represent a confusion matrix of the specified classifier. From the results, we can see, that none of the three classifiers per-formed better, than the linear SVM. The performances are very close, but it is worse at all three classifiers. The qua-dratic reached 0.9715, the polynomial reached 0.9555, the rbf reached 0.9547, while the linear svm was able to reach 0.9766 accuracy.

We tested the K Nearest Neighbor algorithms, based on the number of neighbors and the assignment rule. We used the Euclidian distance to specify the distance of two points. At first, we tested the algorithm with 'nearest'

Table 1 Confusion matrix of classification with Linear SVM.

Linear SVM		Actual	
		1-back	3-back
Recognized	1-back	1160	20
	3-back	40	1340

Table 2 Confusion matrix of classification with Quadratic, Polynomial and RBF SVM.

			Actual	
			1-back	3-back
Quadratic	Recognized	1-back	1160	33
		3-back	40	1327
Polynomial	Recognized	1-back	1152	66
		3-back	48	1294
RBF	Recognized	1-back	1112	28
		3-back	88	1332

assignment rule. This means, that the nearest neighbor decides, when the number of deciding neighbors are equal for the two classes.

The results are shown on Fig. 4. We can see, that the 1NN and 2NN performed the best. They are the same, because in the 2NN case, there is an equality, where the nearest vector decides. With the raising of K, the algorithms performed worse. The confusion matrix of the 1NN classification is shown at Table 3. The accuracy of the 1NN algorithm with nearest rule is 0.7637 (1955/2560). We can see, that this rate is much lower than any from the support vector machines.

In order to raise the rate, we tried K Nearest Neighbor algorithm with 'consensus' rule. At this rule, the K nearest vectors must belong to the same class. If there are vectors from multiple classes, the algorithm will not decide. The results from K = 1 to K = 10 are shown on Fig. 5. We can see, at K = 1 the result is the same as the nearest rule, because at one vector, there is always a consensus. With the raising of the K, the accuracy also raises. From K = 7, it nearly reaches the performance of the support vector machines. But because of the consensus rule, the algorithms do not always classify a vector.

Fig. 6 shows the number of classified vectors depending on the K. Here, we can see, that up to K = 5 there were less than 1000 vectors from 2560 classified. At K = 10, just 538 vectors were classified. If we see the accuracy at the whole dataset, it is just 0.1984 (508/2560). This consensus rule can only work in cases, when we have lot of data, and we have lot of time to decide. In that case, this small number of decisions can be enough.

We tested linear regression on our data set as well. The results are presented in Table 4. The accuracy is very close to the linear svm, it is 0.9742 (2494/2560). The support vector machine with linear kernel classified 6 more vectors correctly than the linear regression.

We tested three different algorithms with various parameters in our study. From the results, we can see, that there are algorithms, that are capable of classifying the vectors, created from physiological data, to infer mental state.

The K Nearest Neighbor algorithm performed best at K = 1. Here it was able to reach 76.37 % accuracy. From the Support Vector Machines, the linear svm performed the best, with 97.66 % accuracy, while the linear regression algorithm reached 97.42 %. With our experiments, we prove, that we are capable for infer the mental state from physiological data, with a high success.
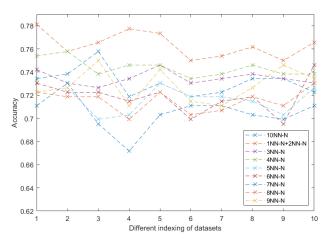

Fig. 4 Performance of the KNN algorithms with nearest rule

Table 3 Confusion matrix of classification with 1NN with nearest rule.

1NN Nearest		Actual	
		1-back	3-back
Recognized	1-back	889	294
	3-back	311	1066

Fig. 5 Performance of the KNN algorithms with consensus rule

Fig. 6 Number of classified vectors based on K

Table 4 Confusion matrix of classification with Linear Regression.

Linear Regression		Actual	
		1-back	3-back
Recognized	1-back	1157	23
	3-back	43	1337

4 Cognitive Games

We can measure four areas of intelligence with cognitive games: Logical-mathematical (1), which measures logic, skill with numbers, abstractions and critical thinking. Verbal-linguistic (2), which relates to facility with words and languages. Musical-rhythmic (3), which provides information about sensitivity to sound music and rhythm. And lastly visual-spatial (4), which deals with spatial judgment and the ability to visualize.

The games measured fall into the following areas of intelligence [37]:

- mathematical-logical:
  - N-back with numbers: numbers are shown to the player, one by one. He has to enter the number n steps before the current one.
  - Pin Code: the player has to remember sequences of numbers displayed for a short time.
  - Counting: the player has to solve adding and subtracting problems using the results from previous calculations.
- verbal-linguistic
  - N-back with characters: similar to N-back with numbers, only with characters.
  - Master Mind: the player has to type in words with same length as the original one. After each one, he can see how many letters were correct and in the correct place, or correct but in a wrong position.
  - Cypher: letters of known short texts are coded with shapes. The player has to decode the characters.
  - Anagram: the player has to create words from the letters presented to him.
- visual-spatial:
  - N-back with shapes: similar to N-back with numbers, only with shapes.
  - Picture Memory: sequences of symbols are shown to the player, which later he has to put in order.
  - Short-term 2D visual memory: the player sees numbered objects on a background. After a short period of time, the objects disappear, and the player has to recall their correct placement in the right order.
    - Space Plane Relation: the player sees a 3D object, and is presented a number of 2D views, from which one is corresponding to the 3D object. He has to pick the correct one.
- musical-rhythmic:
  - N-back with sounds: similar to N-back with numbers, only with sounds.
  - Simon says: the player has to repeat an increasing series of voices and sounds by pressing the correct buttons.

As a more detailed example, the game called N-back with shapes has the following input parameters in Table 5.

The example is a 3-back game with an abstract symbol set. In the first 3 steps, the player cannot guess, since there was no step 3 steps before the current one.

On the 4th step, the player is shown a new picture, but also has to choose the one 3 steps before. This repeats until the full time or the total number of steps is reached.

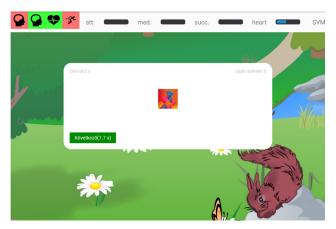At the end of the gameplay, the player is shown his score, and can choose a new game to play.
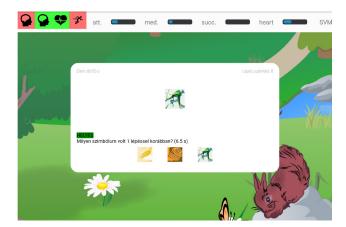

Fig. 7 N-back gameplay beginning

Fig. 8 N-back gameplay in progress

Table 5 Input parameters

Input parameter name	Details
showAnswerCorrect	Option to show if previous answer was correct
showCorrectAnswer	Option to show correct solution after incorrect answer
n	The number of steps since the current guessed symbol
duration	Duration of the game in ms
showDuration	How long disks can be seen in ms
maxStepNumber	Number of total rounds
paletteId	Determines the symbol set
maxSymbol	Size of the symbol set
availableSymbol	The number of symbols from which the player has to identify the correct one

Table 6 Output parameters

Output parameter name	Details
correctAnswers	Number of correctly solved rounds
averageTime	Average round time
correctRation	Percentage of correctly solved rounds
attemptNumber	Number of total attempts

4.1 Cognitive profile

With the help of these cognitive games, we are able to create a cognitive profile for each user. The profile contains an ability number for each of the measurable intelligence fields. This ability number has a similar concept to Intelligence Quotient (IQ), which means that a person's intelligence on a certain field is put on a scale with mean 100 and standard deviation 15. In addition to this, the AdaptEd framework provides the opportunity to create adaptive scales. This means that we can specify filters for the groups of people to put on the scale. For example, there is little use in comparing a person who has dyscalculia with a person who doesn't. We can compare his abilities with people with same disadvantages, or we can specify age groups to make our comparisons more accurate. The cognitive profile does not only store the ability number for the multiple intelligence areas, but for each area, it stores the mental effort classification models separately. Since according to Gardner's MI theory, we can have very different scores on the different fields, it is necessary to measure mental effort for each of them. This part of the profile is stored automatically in the framework and is updated on each new model training. If the last training period was too long ago, the system automatically prompts for a new one to be taken.

The output of the Interest Map method can be integrated to the cognitive profile as well, for more complex and accurate analysis. It creates the opportunity to compare the results of self-evaluation and automated measurements.

5 Using IRT
5.1 Item parameter estimation with previous ability estimates

As described in the previous section, various cognitive games can be used to measure the intelligence of players for different cognitive fields. These measurements in turn can be used to give an estimate for the ability (as used in an IRT model) of a given subject when playing an educational game.

We suppose that a typical educational game focuses on one or two cognitive fields. In this case, we can estimate the ability of a player by weighting the intelligence level of the player for the cognitive fields the game is aiming to develop. However, choosing the correct cognitive fields and weighting ratios are critical.

After this, we can treat the ability of players as known and estimate the parameters of the test items which each corresponds to a task in the educational game.

For this estimation, we first need to choose the type of IRT model we are going to use. The most detailed one is 3PL so we are going to focus on that for the rest of the paper.

First, we need to collect data consisting of player responses for the test items. Let N be the number of players and n be the number of items. Let the N times n-dimensional vector $\boldsymbol{u}$ consist of the responses of the N players for the n items. Let $\boldsymbol{\theta}$ be an N-dimensional vector containing the abilities of players which are already known from their intelligence levels.

The three n-dimensional vectors to be estimated are denoted with $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$, where $\boldsymbol{a}$ contains the discrimination, $\boldsymbol{b}$ contains the difficulty and $\boldsymbol{c}$ contains the guessing parameter for each item.

Maximum likelihood estimation is one of the methods which can be used in this case. Using MLE, the likelihood function $L(\boldsymbol{u} \mid \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ can be expressed as:

$$L\left(\boldsymbol{u} \mid \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\right) = \prod_{i=1}^{n} P_i^{u_i} Q_i^{1-u_i}, \tag{2}$$

where P_i is the probability function which gives the probability that a player answers correctly for item i and $Q_i = 1 - P_i$.

As it is standard when using MLE, the logarithm of the likelihood function can be used for maximization since it will take its maximum at the same point as the likelihood function. Due to the nature of the function, numerical procedures are to be used for this task, such as the Newton-Raphson procedure.

The method described above simplifies the estimation of item parameters. As opposed to estimating subject and item parameters simultaneously, in this case, the likelihood function only depends on the item parameters.

5.2 Extending IRT models with mental effort
5.2.1 Extending the model
Current IRT models do not take the mental effort of subjects into account when estimating the probability of solving a given item correctly. In this section, we propose a modification of simple IRT models which uses an extra parameter for subjects. This parameter, denoted with δ, describes the mental effort put by the given subject when solving the item.

When we add this extra parameter, the modified item response function for 3PL becomes the following:

$$p(\theta, \delta) = c + \frac{\delta - c}{1 + e^{-a(\theta - b)}}. \tag{3}$$

This is the same formula as which is used for the four-parameter logistic model (4PL), but in this case δ is considered to be a parameter of the subject rather than an item parameter. This means that the value of the parameter δ influences the upper asymptote of the ICC (which in turn means it cannot be greater than 1). Thus, a transformation of the value describing the mental effort of a subject can be used when we substitute in the formula above. Intuitively, this means that if subjects are in a bad mental state (fatigued, etc.), they will solve items correctly with lower probability.

When estimating these modified IRT models, the process could get difficult because the model is no longer unidimensional: the subjects now have multiple parameter values and the new parameter for a subject also varies from task to task. Because of this, we assume that the δ parameter is averaged out when a given subject solves a large number of items. We also assume that the value of this parameter does not vary considerably when the same subject solves different tasks and the average value is also more or less the same for different subjects. This can be guaranteed if we choose the function which transforms the mental effort value into the δ value appropriately. Due to these assumptions, we substitute δ with a constant value when estimating the item parameters of these models.

5.2.2 Validating the extended model
In order to check the validity of the IRT model extended with the mental effort parameter, we used simulation and took measurements of the performance and physiological signals (EEG, ECG) of subjects while playing a game consisting of IRT test items. The simulation was needed to estimate the number of subjects needed for the measurement, as well as to validate the convergence of the model.

For the measurements, we used the cognitive game named *Counting* which is introduced in Section 4. This game is available in both web and mobile environments, and can be used to measure mathematical-logical ability. It contains simple arithmetic calculations: addition and subtraction, the first operand of the operation always being the result of the previous one. Sometimes, abstract shapes take the place of numbers. In this case, the values of such shapes are always given to the player first, then they have to recall the values several steps later. If the player gives a wrong answer to a question, the next question will "restart": both operands will be given once again with the values of shapes reset. The operands and results of the arithmetic operations are always nonnegative integers selected from a pre-determined, fixed interval (e.g. 0 to 99). Players have to solve the tasks within a given time limit, without using any form of help (calculator, sheets of paper, etc.). The aim is to maximize the number of calculations solved while minimizing the number of mistakes.

The game described above corresponds to a test using the extended IRT model which includes mental effort. The ability measured corresponds to the mathematical-logical ability of the subject, while the mental effort can be calculated from the signals provided by the sensors described in Section 2.1. Each test item corresponds to an arithmetic operation (addition or subtraction). To simplify, we take multiple assumptions. First, each test item is assumed to have an unknown but equal difficulty (*b*) parameter, which holds only approximately: it is possible that an operation with two particular operands is more difficult than another operation with different operands. Similar to the 2PL model, the discrimination (*a*) parameter of each item is assumed to be 1. The guessing (*c*) parameter is approximated from the reciprocal of the number of possible answers which can be calculated by adding 1 to the length of the interval the result is taken from. The length of this interval is known to the player.

These simplifications have multiple reasons: first, it is often advisable to solve a relaxed problem first, where the original problem is replaced by an approximation which is easier to solve. Moreover, in order to estimate item

parameters, the algorithms described in Section 2.2.2 need data from hundreds of subjects which was not available for the game we used.

The mapping described above is summarized in Table 7.

5.2.3 Simulating the extended model

To analyze the model before measurement, we created a simple simulation algorithm implemented in Python. We examined the amount of measurements needed to achieve results with relatively little noise and the effect of the fine tuning of the parameters on the model. The simulator uses the modified, bivariate item response function described in Section 5.2.1.

The simulator also takes multiple assumptions. First and foremost, the simulated subjects give correct answers with the exact probability described in the IRT model. Their ability parameter is a random, normally distributed value, which corresponds to reality quite well. The mental effort of simulated subjects is characterized using the δ parameter which can be an arbitrary function of elapsed time. To simplify things, we used a decreasing linear function, which means that the probability of the correct answer is greater in the beginning than at the end. When simulating an answer, the time needed is taken from an exponential distribution for which the λ parameter is equal for each subject which means each subject will have the same average response time: the reciprocal of λ. All item parameters (difficulty, discrimination, guessing) are the same for each simulated item.

Next, we describe some simulation passes with the same fixed parameter values. The number of simulated subjects is increased tenfold for each pass from 1 to 1000 which means there are four passes with 1, 10, 100 and 1000 subjects, respectively. The parameter values are given in Table 8.

In each simulation pass, we looked at the averages of the data for the simulated subjects. For each pass, we examined the number of total and correctly solved test items and the percentage of correct answers summed over a moving, 3-minute period.

Table 7 Mapping the *Counting* game to an extended IRT model

ability	mathematical-logical
mental effort	calculated from EEG and ECG
test items	arithmetic operations
difficulty	equal
discrimination	equal (1)
guessing	1 / (length of interval + 1)

When simulating only one subject, the results are very noisy, but a decreasing trend for the performance is visible. Increasing the number of subjects and taking the average of data for each subject, the graphs are smoothening gradually, revealing the linear decrease we used, as seen in Fig. 9 and Fig. 10.

Table 8 Example of parameter values for simulation

subject ability	normal distribution, $\mu = 0$, $\sigma = 1$
subject δ	decreasing from 0.95 to 0.75
subject response time	exponential distribution, $\lambda = 1 / 3$
item difficulty	−3
item discrimination	1
item guessing	1 / 100
simulated timespan	40 minutes

(a) 10 subjects

(b) 100 subjects

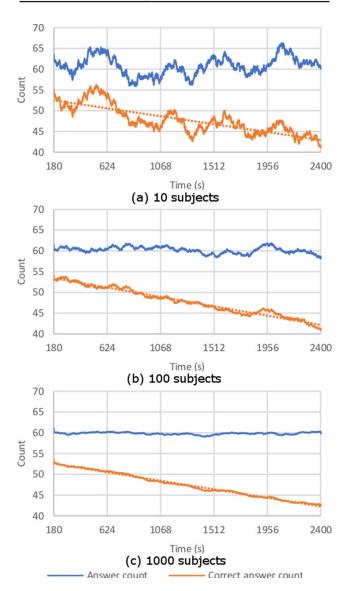(c) 1000 subjects

— Answer count   — Correct answer count

Fig. 9 Averages of answer count and correct answer count for different numbers of simulated subjects summed over a 3-minute moving period
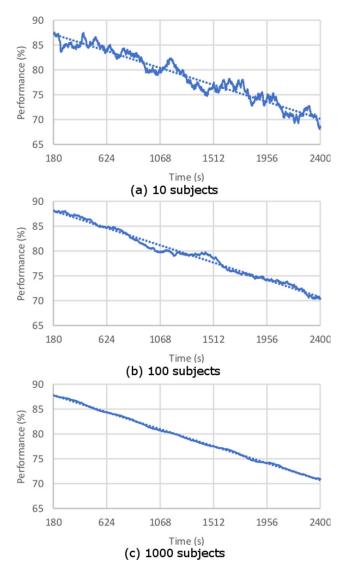
Fig. 10 Averages of the performance for different numbers of simulated subjects taken over a 3-minute moving period
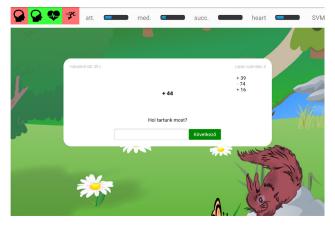

Fig. 11 Averages of the performance for different numbers of simulated subjects taken over a 3-minute moving period

The simulation results show the convergence of the model, which means the shape of the function according to the performance varies over time becomes gradually more clearly visible when the number of subjects is increased. The results also show that we should be able to obtain reasonably smooth data with the measurement of approximately 100 subjects.
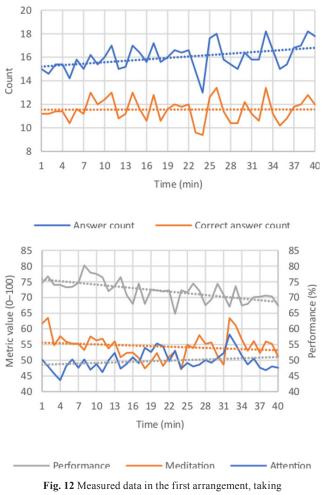
5.2.4 Measurement of the extended model

11 adults (10 males, 1 female) aged 20 to 27 participated in the measurements. Every participant reported healthy with no diagnosed cardiovascular and neurological disorders. Each participant was asked not to take any stimulants (caffeine, cigarettes, alcohol, drugs, etc.) in the 8 hours preceding the measurement. Each measurement began with a two-minute relaxation phase.

We conducted the measurements in two different arrangements. Both arrangements used the cognitive game *Counting* as the basis of the extended IRT model, as described in Section 5.2.2. Furthermore, the input parameters of the generated questions were the same for both arrangements: the minimum and maximum accepted values for each operation were 0 and 99, respectively, the maximum number of symbols to remember at a time was 2, and the probability of a symbol appearing in a question was 25 %.

In the first arrangement, the measurement consisted of 40 1 minute phases with short breaks in between. In each phase, we measured the number of *total* and *correct answers* and the percentage of correct answers (*performance*). To measure mental effort, *NeuroSky MindWave*, a one-channel EEG headset was used. This device allowed us to measure two performance metrics calculated based on raw EEG signals: *meditation* and *attention*. Each value was outputted by the device as an integer ranging from 0 to 100. We sampled the output every second, and then took the average of the values for each 1 minute phase. 5 out of 11 subjects participated in this arrangement.

In the second arrangement, the measurement lasted 40 minutes, uninterrupted. To measure cognitive performance, we recorded every answer and summed the number of *total* and *correct answers* over a 5 minute

Table 9 Input parameters for both arrangements

Minimum result	0
Maximum result	99
Number of symbols	2
Symbol probability	25 %

Fig. 12 Measured data in the first arrangement, taking the average of 5 subjects

moving period. We also calculated the percentage of correct answers over the same period (*performance*). Mental effort was measured using two devices in this arrangement. *Emotiv Insight*, a five channel EEG headset provided six performance metrics calculated every second from raw EEG signals: *interest, stress, focus, engagement, relaxation and excitement*. Each of these values were outputted by the device as a floating point number between 0 and 1. ECG signals were also recorded *using Zephyr HxM BT*, a wireless heart rate monitor. From the raw heart rate values, we calculated *heart rate variability* using two methods: SDNN (standard deviation of peak-to-peak intervals) and Burg's algorithm. Each signal was averaged over the same 5 minute moving period as described above. The second arrangement was used on 6 out of 11 subjects.

When using the data from the first arrangement, the average of the values was taken from all measurements. The data shows an upward trend in the number of total answers with the number of correct answers remaining more or less constant, which means the percentage of correct answers
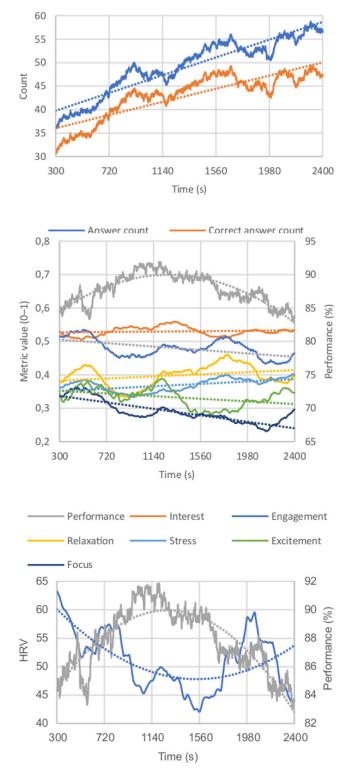

Fig. 13 Measured data in the second arrangement, using a 5 minute moving period and taking the average of 6 subjects

decreases over time. The EEG performance metrics showed that meditation decreased over time with a slight increase in attention. It should be noted that both the answer count and EEG metrics smoothened when more measurements

were averaged, which is another indication of convergence and that the EEG metrics can be used as the basis of the IRT model's δ parameter, however, more measurements are needed to determine their exact relationship.

When evaluating the results from the second arrangement, we also took the average of values from all measurements. The data shows the performance percentage first increases, then gradually decreases over time. Multiple explanations can be given for this phenomenon. For example, when new participants first encounter the game, they need some time to familiarize themselves with the game mechanics. Thus, an approximately 10–15 minutes long learning phase should be incorporated before taking a measurement. An increasing trend can be seen on both the number of total and correct answers with the number of correct answers increasing more slowly, causing the percentage of correct answers to decrease. The faster increase in the beginning can be attributed to the player "getting the hang" of the game, as before. However, the answer count is still increasing at the end, perhaps indicating the growing impatience of the player: the idea of thinking less over a difficult question to increase the number of questions solved. The evaluation of the data provided by the EEG device shows that the interest value is averaging a constant value over time, with stress and relaxation slightly increasing. Excitement, engagement and focus show a more prominent decreasing trend, which can be attributed to getting bored with the repetitive task (in case of excitement) and having fatigue (in case of engagement and focus). The latter two also seem to be strongly correlated. The HRV values show a strong negative correlation with performance, lower values indicating better performance. The same statements can be made after evaluating the second arrangement: the metrics smoothened when taking averages of measurement, which means the model converges, and that the metrics can be used for the calculation of δ, the exact nature of the relationship can again only be determined from more measurements. However, the three decreasing EEG values, especially focus, as well as the HRV value can all be strong indicators to the value of δ.

6 Conclusion

In this paper, we summarized our work regarding new and innovative methods used in education. We reviewed the main reasons why players need adaptive solutions when playing serious games. We presented a solution for adaptive gaming which utilizes a wide range of techniques from a variety of fields, including, but not limited to computer science, cognitive psychology, psychometrics and mathematics.

We created a cognitive profile, which automatically measures and stores information about the most important cognitive abilities of the user. Both mental effort data and intelligence measurements are available for analysis as part of the profile. The numbers describing abilities are used as an input parameter for the IRT model.

We also discussed the fundamentals of IRT models, including the methods which can be used to estimate these models. Additionally, we described a new way to estimate item parameters in IRT models in the case where the parameters of subjects are already known. We proposed an extension of the three parameter IRT model which is able to take mental state into account. We validated our new model using simulation and by conducting measurements using commercially available EEG and ECG devices. Both methods showed the convergence of the model.

In this paper, we showed that we are capable to infer the mental state from physiological data. With the use of classification algorithms, we can reach up to 97 % accuracy, when classifying data vectors to different mental states. With the use of this information, which has key role in the performance of the user during serious games, we can create an IRT model, which takes into account the current physiological state of the user as the fourth parameter.

As a combined result, a multifaceted and extensible system was created, which can also be a starting point and tool for future researches.

References

[1] Michael, D. R., Chen, S. L. "Serious games: Games That Educate, Train, and Inform", 1st ed., Muska & Lipman/Premier-Trade, 2005.

[2] Sík Lányi, C., Brown, D. J., Standen, P., Lewis, J., Butkute, V. "Results of User Interface Evaluation of Serious Games for Students with Intellectual Disability", Acta Polytechnica Hungarica, 9(1), pp. 225–245, 2012. [online] Available at: https://uni-obuda.hu/journal/Sik-Lanyi_Brown_Standen_Lewis_Butkute_33.pdf [Accessed: 1st March 2018].

[3] Csikszentmihalyi, M. "Beyond Boredom and Anxiety", 1st ed., Jossey-Bass Publishers, San Francisco, USA, 2000.

[4] Sampayo-Vargas, S., Cope, C. J., He, Z., Byrne, G. J. "The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game", Computers & Education, 69, pp. 452–462, 2013.
https://doi.org/10.1016/j.compedu.2013.07.004

[5] Hambleton, R. K., Swaminathan, H. "Assumptions of Item Response Theory", In: Item Response Theory: Principles and Applications, 1st ed., Springer, Dordrecht, Netherlands, 1985, pp. 15–31.
https://doi.org/10.1007/978-94-017-1988-9_2

[6] Hambleton, R. K., Swaminathan, H. "Estimation of Ability", In: Item Response Theory: Principles and Applications, 1st ed., Springer, Dordrecht, Netherlands, 1985, pp. 75–99.
https://doi.org/10.1007/978-94-017-1988-9_5

[7] Hambleton, R. K., Swaminathan, H. "Estimation of Item and Ability Parameters", In: Item Response Theory: Principles and Applications, 1st ed., Springer, Dordrecht, Netherlands, 1985, pp. 125–150.
https://doi.org/10.1007/978-94-017-1988-9

[8] Hambleton, R. K., Swaminathan, H., Rogers, H. J. "Fundamentals of Item Response Theory", Sage Publications, Newbury Park, California, USA, 1991.

[9] Bock, R. D., Moustaki, I. "15 Item Response Theory in a General Framework", In: Rao, C. R., Sinharay, S. (eds.), Handbook of Statistics: Psychometrics, Vol. 26, Elsevier, Amsterdam, Netherlands, 2007, pp. 469–513.
https://doi.org/10.1016/S0169-7161(06)26015-2

[10] ChuDuc, H., NguyenPhan, K., NguyenViet, D. "A Review of Heart Rate Variability and its Applications", APCBEE Procedia, 7, pp. 80–85, 2013.
https://doi.org/10.1016/j.apcbee.2013.08.016

[11] Berntson, G. G., Bigger, J. T., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., Stone, P. H., van der Molen, M. W. "Heart rate variability: Origins, methods, and interpretive caveats", Psychophysiology, 34(6), pp. 623–648, 1997.
https://doi.org/10.1111/j.1469-8986.1997.tb02140.x

[12] Fairclough, S. H., Mulder, L. J. M. "Psychophysiological processes of mental effort investment", In: Wright, R. A., Gendolla, G. H. E. (eds.), How motivation affects cardiovascular response: Mechanisms and applications, American Psychological Association, Washington, DC, USA, 2012, pp. 61–76.
https://doi.org/10.1037/13090-003

[13] Pagani, M., Lombardi, F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., Sandrone, G., Malfatto, G., Dell'Orto, S., Piccaluga, E. "Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog", Circulation Research, 59(2), pp. 178–193, 1986.
https://doi.org/10.1161/01.res.59.2.178

[14] Schomer, D. L., Lopes da Silva, F. H. "Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields", 6th ed., Lippincott Williams & Wilkins, Philadelphia, Pennsylvania, USA, 2012.

[15] Lee, J. C., Tan, D. S. "Using a low-cost electroencephalograph for task classification in HCI research", In: Proceedings of the 19th annual ACM symposium on User interface software and technology, Montreux, Switzerland, 2006, pp. 81–90.
https://doi.org/10.1145/1166253.1166268

[16] Wang, Q., Sourina, O., Nguyen, M. K. "EEG-Based "Serious" Games Design for Medical Applications", In: 2010 International Conference on Cyberworlds, Singapore, Singapore, 2010, pp. 270–276.
https://doi.org/10.1109/cw.2010.56

[17] Dimitriadis, S. I., Laskaris, N. A., Tsirka, V., Vourkas, M., Micheloyannis, S. "What does delta band tell us about cognitive processes: A mental calculation study", Neuroscience Letters, 483(1), pp. 11–15, 2010.
https://doi.org/10.1016/j.neulet.2010.07.034

[18] Grabner, R. H., Fink, A., Stipacek, A., Neuper, C., Neubauer, A. C. "Intelligence and working memory systems: evidence of neural effciency in alpha band ERD", Cognitive Brain Research, 20(2), pp. 212–225, 2004.
https://doi.org/10.1016/j.cogbrainres.2004.02.010

[19] Nering, M. L., Ostini, R. (eds.) "Handbook of Polytomous Item Response Theory Models", 1st ed., Taylor & Francis Group, New York, USA, 2010.

[20] Thissen, D., Orlando, M. "Item response theory for items scored in two categories", In: Thissen, D., Wainer, H. (eds.), Test Scoring, Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, USA, 2001, pp. 73–140.

[21] Rasch, G. "Probabilistic models for some intelligence and achievement tests", Danish Institute for Educational Research, Copenhagen, Denmark, 1960.

[22] Hambleton, R. K., Jones, R. W. "Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development", Educational Measurement: Issues and Practice, 12(3), pp. 38–47, 1993.

[23] Bock, R. D., Aitkin, M. "Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm", Psychometrika. 46(4), pp. 443–459, 1981.
https://doi.org/10.1007/bf02293801

[24] Han, K. T. "Fixing the c Parameter in the Three-Parameter Logistic Model", Practical Assessment, Research & Evaluation, 17(1), pp. 1–24, 2012.

[25] Gardner, H. E. "Multiple Intelligences: The Theory in Practice", Basic Books, New York, USA, 1993.

[26] Gardner, H. E. "Intelligence Reframed: Multiple Intelligences for the 21st Century", Perseus Books Group, New York, USA, 2000.

[27] Gyarmathy, E., Herskovits, M. "Képességek vizsgálata az érdeklődés térképének segítségével" (Investigation of abilities with the help of mapping interests), Pszichológia, 19(4), pp. 437–458, 1999. (in Hungarian)

[28] Gyarmathy, E. "The Gifted and Gifted Education in Hungary", Journal for the Education of the Gifted, 36(1), pp. 19–43, 2013.
https://doi.org/10.1177/0162353212471587

[29] Tirri, K., Nokelainen, P., Komulainen, E. "Multiple Intelligences: Can they be measured?", Psychological Test and Assessment Modeling, 55(4), pp. 438–461, 2013.

[30] Mazaheri, S., Fatemi, O. "An adaptive e-learning system based on Gardner's eight intelligence theory", In: 2014 6th Conference on Information and Knowledge Technology (IKT), Shahrood, Iran, 2014, pp. 128–133.
https://doi.org/10.1109/ikt.2014.7030346

[31] Kelly, D., Tangney, B. "A Framework for using Multiple Intelligences in an ITS", In: Lassner, D., McNaught, C. (eds.) Proceedings of ED-MEDIA 2003 - World Conference on Educational Multimedia, Hypermedia & Telecommunications, Honolulu, Hawaii, USA, pp. 2423–2430, 2003.

[32] Hafidi, M., Lamia, M. "A personalized adaptive e-learning system based on learner's feedback and learner's multiple intelligences", In: 2015 12th International Symposium on Programming and Systems (ISPS), Algiers, Algeria, 2015, pp. 1–6.
https://doi.org/10.1109/isps.2015.7244969

[33] Denig, S. J. "Multiple Intelligences and Learning Styles: Two Complementary Dimensions", Teachers College Record, 106(1), pp. 96–111, 2004.
https://doi.org/10.1111/j.1467-9620.2004.00322.x

[34] Forstner, B., Szegletes, L., Angeli, R., Fekete, A. "A general framework for innovative mobile biofeedback based educational games", In: 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), Budapest, Hungary, 2013, pp. 775–778.
https://doi.org/10.1109/coginfocom.2013.6719204

[35] Szegletes, L., Forstner, B. "Reusable framework for the development of adaptive games", In: 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfocom), Budapest, Hungary, 2013, pp. 601–606.
https://doi.org/10.1109/coginfocom.2013.6719173

[36] Szegletes, L., Köles, M., Forstner, B. "The design of a biofeedback framework for dynamic difficulty adjustment in games", In: 2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom), Vietri sul Mare, Italy, 2014, pp. 295–299.
https://doi.org/10.1109/coginfocom.2014.7020465

[37] Pomázi, K., Gazdi, L., Radostyán, B., Szabó, M., Szegletes, L., Forstner, B. "Self-standardizing cognitive profile based on gardner's multiple intelligence theory", In: 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Wroclaw, Poland, 2016, pp. 317–322.

[38] Bodolai, D., Gazdi, L., Forstner, B., Szegletes, L. "Supervising Biofeedback-based serious games", In: 2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Gyor, Hungary, 2015, pp. 273–278.
https://doi.org/10.1109/coginfocom.2015.7390603

[39] Kirchner, W. K. "Age differences in short-term retention of rapidly changing information", Journal of Experimental Psychology, 55(4), pp. 352–358, 1958.
https://doi.org/10.1037/h0043688