M-LEARNING PLATFORM FOR ASSESSMENT AND PERSONALIZED LEARNING OF THAI LANGUAGE BY PRIMARY SCHOOL CHILDREN

SAOWALAK RUNGRAT

A Thesis Submitted to the Graduate School of Naresuan University
in Partial Fulfillment of the Requirements
for the Master of Science in (Computer Science)
2020

M-LEARNING PLATFORM FOR ASSESSMENT AND PERSONALIZED LEARNING OF THAI LANGUAGE BY PRIMARY SCHOOL CHILDREN

SAOWALAK RUNGRAT

A Thesis Submitted to the Graduate School of Naresuan University
in Partial Fulfillment of the Requirements
for the Master of Science in (Computer Science)
2020

Thesis entitled "M-Learning platform for assessment and personalized learning of Thai language by primary school children"

By SAOWALAK RUNGRAT

has been approved by the Graduate School as partial fulfillment of the requirements for the Master of Science in Computer Science of Naresuan University

Oral Defense Committee

.... Chair
(Associate Professor Pensiri Tongpadungrod, Ph.D.)

.... Advisor
(Assistant Professor Antony James harfield, Ph.D.)

.... Co Advisor
(Supiya Charoensiriwath, Ph.D.)

.... Internal Examiner
(Assistant Professor Jaratsri Rungrattanaubol, Ph.D.)

.... Internal Examiner
(Wansuree Masagram, Ph.D.)

Approved

....
(Professor Paisarn Muneesawang, Ph.D.)

Dean of the Graduate School

Title	M-LEARNING PLATFORM FOR ASSESSMENT AND PERSONALIZED LEARNING OF THAI LANGUAGE BY PRIMARY SCHOOL CHILDREN
Author	SAOWALAK RUNGRAT
Advisor	Assistant Professor Antony James harfield, Ph.D.
Co-Advisor	Supiya Charoensiriwath, Ph.D.
Academic Paper	Thesis M.S. in Computer Science, Naresuan University, 2020
Keywords	Personalized learning, Item response theory, Intelligent tutoring systems, M-learning, Language learning, KidLearn

ABSTRACT

KidLearn is an M-learning platform for primary school children with a personalized learning component for Thai language learning. The purpose of this study is to describe and evaluate the personalization algorithm. By applying item response theory, the algorithm calculates ability in language topics based on responses to test questions and selects new content aimed at maximizing each child's improvement in ability. An experiment was undertaken in 3 schools with 47 children with low-ability or learning difficulties in Thai language reading. The results show that improvements in the children's ability in each topic were highly correlated with the ability calculated by the personalization algorithm. Therefore, as well as KidLearn providing an efficient means for boosting a child's language learning across different topics, it effectively predicts a child's language ability which provides educators an unobtrusive testing tool for monitoring progress.

ACKNOWLEDGEMENTS

The authors would like to express thanks to Issarapa Chunsuwan, Kanokporn Vibulpatanavong, and Nichara Ruangdaraganon for guiding the development of KidLearn, and to Samakkhi Rat Bamrung School, Pathumthani Municipality Secondary School and Pongsuwanwittaya School for testing the KidLearn tutoring system. This project was funded by Thailand Graduate Institute of Science and Technology (TGIST) programmed under National Science and Technology Development Agency (NSTDA).

SAOWALAK RUNGRAT

TABLE OF CONTENTS

Page

1 INTRODUCTION

The number of cases of children with learning disabilities is rising in Thailand. Children who are “left behind” due to these difficulties are likely to develop further problems because not only they are learning slower than other children but often they do not learn at all due to lack of confidence. This can lead to personal problems such as inappropriate behavior in the classroom or isolation from social interaction with others.

Learning disabilities can be divided into three categories: reading disorders, writing disorders, and mathematic disorders [1]. The reading disorder (dyslexia) is the most common at around 80 percent of children with learning disabilities [2]. A child with a reading disorder will have problems with analyzing sound, remembering letters, associating letters with their sound, reading speed, and ability to recognize words. These are essential skills for children for the development of learning in other fields.

Cases of children with learning disabilities requires support from doctors, special teachers, and in some cases special education schools. Often children need individual help because each child can have different problems and their reading is at different levels. The increase in cases of children with learning disabilities means that there are not enough qualified teachers to give the individual support that each child needs. The aim of this research is to investigate technological solutions that can provide support in the assessment and training of children with reading disorders.

A number of tools already exist for assessing children for learning disabilities alongside the diagnosis and treatment by a doctor. These tools, which have been developed both in Thailand and abroad, cover similar learning disorder topics but are different in language and approaches. In English language learning for example, there are a wider range of non-technological tools (e.g. Learning disabilities self-screening Tool) and technological tools (e.g. “EasyLexia”, a mobile app for children to improve some of their fundamental reading skills [3]). Thailand does not have as many widespread tools to test these basic skills. Traditionally, the assessment of children with reading disorders was tested using paper. In Thailand, one such assessment was digitalized using a web application [4]. The assessment is easily digitalized because it often has a limited number of tests which are all conducted in the same way. However, training is more complicated because it should take into account the child’s current level and their progress throughout their interaction with the system. Therefore, this research utilizes a form of personalized learning to deliver exercises to each child that are suitable for their current learning level. The purpose of personalizing the learning is so that children can practice until they reach their goals, and gradually develop at their own pace.

Personalized learning is a broad term covering approaches that tailor education to individual learners. Personalized implies that the learning objectives, instructional strategies and instructional content are selected based on a child's ability. Personalized learning aims to solve the problem that the pace, style and content of education in classrooms or group-based learning is not be aligned with each individual person. While it is feasible for personalized learning to occur without technology, typically the benefits of availability and scalability mean that technological solutions are most appropriate for implementing personalized learning. Within the area of personalized learning, researchers have defined types such as adaptive learning, individualized learning, differentiated learning, and competency-based learning [5]. In this project, the focus is on adaptive learning.

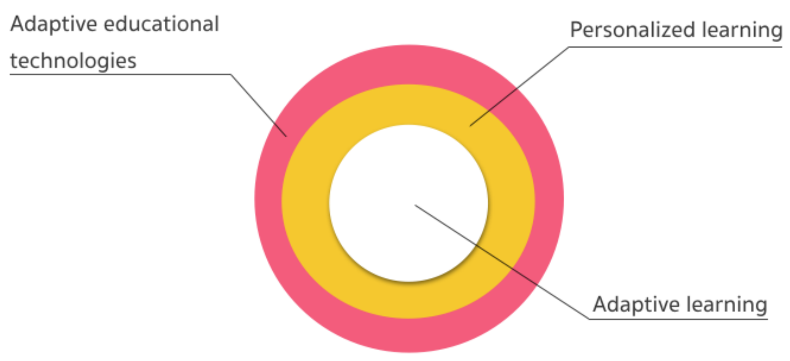


Figure 1. Personalized learning level

Adaptive learning is one of the techniques in providing personalized learning (as Figure 1) with the aim of providing effective, customized learning pathways to individual students. Adaptive learning systems gather data about the learner from interactions within the system itself. For example, the system might record the amount of time spent on reading material and the responses to quiz questions, and then adapt the pathway through the future learning content. Adaptive learning systems use algorithms for evaluating results from student responses and interaction, and then adjusting teacher and media interventions to deliver appropriate learning content to learners [6]. There are three general types of adaptive learning systems: closed, open and hybrid. They vary in terms of control, configuration, time-to-use, refinement and content sorting. A closed system comes with a fixed curriculum ready for the learner and they follow the path provided by the system. An open system gives the user control of the learning pathway and can be configured to their preferences. Many systems are hybrid open-closed systems which enable some limited control or configuration. Adaptive learning modules are designed to provide content, sorting content and evaluations level to maximize their learning outcomes based on their current ability [7].

An intelligent tutoring system (ITS) is a computerized tool that delivers customized content to learners, often without any human intervention [8]. Such systems are popular because they offer the learning opportunities "anytime, anywhere" and therefore make learning more accessible to a wider audience. Typically, an ITS can employ adaptive learning techniques to offer a personalized learning experience to the learner.

When technology is applied to personalized learning, it should provide learners with a uniquely tailored learning path as though each learner has the attention of an individual expert. By collecting data on the learner’s past activities and interactions, recommender algorithms suggest lessons, feedback, and assessments that best match the learner’s ability and enable them to overcome their weaknesses [9]. Personalized learning’s key benefits include: a) improving learning outcomes and learning experience, b) supporting a more active approach to teaching and c) enabling learning at scale in a sustainable and cost-effective way. A key ingredient of personalized learning is a system for effectively evaluating the learner's ability and recommending suitable learning content [10].

Computerized Adaptive Testing (CAT) is used to precisely evaluate an examinee’s ability by providing a tailored path through a bank of test items. A key ingredient is the algorithm that selects the most appropriate test item based on the examinee’s ability [11]. Both CAT and PL share a need to effectively evaluate the ability of participants. Where they differ is that CAT recommends test items with the purpose of obtaining a more accurate estimate of the ability, while PL recommends interventions with the purpose of increasing the ability.

Several researchers have noticed the similarity and adapted CAT techniques to PL. The most popular technique, Item Response Theory (IRT), is a statistical measurement model to determine a test taker's ability and their probability of answering a given question correctly [12]. Within the field of IRT, there are two main categories of models: a) dichotomous models which have two types of responses (e.g. correct or incorrect), and b) polytomous models which have multiple types of responses. Given sufficient assessment data (the “responses”), an IRT model is applied to obtain the difficulty parameter and the discrimination parameter for every question (or “item”). For a given test taker, these parameters together with the responses for other questions are sufficient to predict the probability that the test taker will respond to the item correctly. The technique enables assessment systems that dynamically select items to maximize the information about the ability of the test taker and to end the test when the system can predict the test taker’s answer above a given confidence threshold. IRT can significantly reduce the length of assessments by up to 50% [13]. While IRT is highly popular in commercial CAT products, examples of IRT applied to personalized learning are relatively rare and are not yet found in commercial products. The research into IRT for personalized learning can be grouped into two broad categories: assessment-focused and training-focused.

In the area of primary education, the eDia system [14] is an e-learning assessment platform covering reading, mathematics and science used in a large number of schools in Hungary for a number of years. IRT is applied to establish formative assessments that enable diagnostics and improvements in teaching. Results of the long-term study suggest IRT and the platform supports adjusting teaching and learning processes to the individual needs of students.

Within higher education, numerous studies have utilised IRT in assessments of computer science related subjects, such as the adaptive assessment for introductory programming [15]. Typically, such systems can recommend appropriate problems based on student ability. In a study by Yacob et al. [16], undergraduate students in a programming course experienced different learning paths through multiple choice problems that were adapted based on item difficulty and learner ability. The system filtered unsuitable course materials for students, and also helped identifying the items most likely needing for modification by teachers. Kustiyahningsih et al. conducted studies on the use of CAT based on IRT in E-Learning systems [17]. Two groups of 88 students undertook a series of assessments where one group was served all questions non-adaptively and the other group was served each question based on their responses to previous items based on IRT. The algorithm employed a strategy of increasing the difficulty level of the questions in response to correctly answered questions. The results suggest that the adaptive group showed a greater increase in ability compared to the non-adaptive group.

In the second category of related work, there are several examples of IRT applied directly to training or learning. A suitable example is how IRT can be used for vocabulary practice as proposed by Chen & Chung [18] in their work on a mobile-based e-learning system for higher education students studying English as a foreign language. The proposed algorithm chooses a suitable strategy for extending or shortening the memory cycle activities based on the ability of students and the difficulty of the content. The work is similar to the current study, except that IRT is applied for a linear outcome of more/less memory cycle activities and target learners are university students.

The motivation for the author is to explore opportunities to support children with learning difficulties in Thai language using personalized learning. The main contribution of the thesis is:

1. the development of the KidLearn mobile application and backend system for training students with learning difficulties in Thai language
2. an algorithm for selecting content based on Item Response Theory as a new technique for personalized learning

The research question considered is: **Can IRT be effectively applied to personalized learning of language skills for primary school children with learning difficulties?**

The thesis is derived mainly from the results of two published research papers [19, 20]. It is organized into a literature review of IRT-related assessment and training research, a methodology section describing the IRT-inspired algorithm for personalized learning and the development of the KidLearn application and platform, and a results section with the evidence for the effectiveness of the platform and algorithm.

2 LITERATURE REVIEW

The context for the current study is the assessment of reading skills in early years or primary school children. In Thailand, as well as abroad, reading disorders are the highest contributors to learning disabilities [21]. As mentioned earlier, reading disorders include problems analyzing sounds, remembering letters, associating letters with their sounds, reading speed, and ability to recognize words. In this section, existing tools and studies undertaken relating to reading disorders are introduced.

Research into the application of mobile apps for children with learning disabilities suggests there are promising prospects. In one study [3], an application called EasyLexia, consisting of a series of games covering reading, writing and mathematics, was evaluated. The results of the evaluation highlighted that children were able to demonstrate progress over a short period of time depending on their level of dyslexia.

For Thai language, there is a reading test proposed by Vibulpatanavong for screening for learning difficulties [22]. The assessment consists of 5 types of questions: non-word, word, letter, blending, and segmenting. In another study, "Rama Pre-Read: RPR" was proposed for evaluating reading skills [23]. The test topics are as follows: Initial sound matching (ISM), Letter naming (LN), Rapid letter naming (RLN), Letter sound (LS) and Naming (CN). This test will test 4 to 5-year old and be randomly selected for testing in all schools, both public and private the 18 schools. The initial sound matching test was 45 percent and letter naming was 68 percent. The result was a study and family income. Include reading activities translation may not be accurate. The remaining topics are voice writing and category naming. Topics are research abroad, it's a skill that can predict the readiness as well. It may be used as a teaching exercise for children. This assessment is tested not for voice or speech, which will mislead children. These sources informed the development (by the authors) of an assessment for Thai reading skills administered by iPad app [24]. The application is concerned only with assessment and offers no support for training children.

Research for training Thai language reading skill of iPad is KengThai [25]. KengThai is application training Thai language skill have functions write read and games. The app can be connected Facebook for sheare and rank score. The application has many functions interesting for my project, but my application can help adaptive learning for each child combine from doctor or expert.

The example research about development of adaptive learning from personal data and comparison with non-adaptive learning [26]. The research is analyzing the factors of learning patterns and learning behaviors of students. The learning pattern is divided into Sequential Processing Skills, Discrimination, Analytic Skill, and Spatial Skill. The Learning behavior is divided into Learning Achievement Student, Learning effectiveness and the concentration Degree. The results of adaptive learning adaptation test will make children more effective in learning than in non-adaptive learning. This is because some of the measurements are difficult because the adjustments in this exercise are used in mathematics. This research can only be used in some analytical.

Within research use IRT is a technique that can significantly reduce the length of assessments [13] mentioned earlier. There are few examples of IRT applied to training. An example applying IRT application of Componential IRT (2PL) model for diagnostic test in a standard-conformant e-learning system called “IDEAL” [27]. It has been said that IRT algorithm that has been shown very effective in estimating a learner's latent ability. In the system using dichotomous responses (true; 1/false; 0) to learn web-based learning content management system that was developed based on XML technology. In experiment two test the programming course C++ (49 questions) and XML (46 questions) and the result will model is investigated using Chi-Square test. After the response the system rescale each difficulty parameter and estimate ability based on current item. The system will stop when the training no improvement on difference between current and the previous. The final system result, where an IRT model diagnosis and remediation is implemented, shows that the proposed approach is effective.

One example is a personalized curriculum sequencing utilizing modified item response theory for web-based instruction [28] is a research on adaptive learning on websites that can be adapted to the needs of students. It is a hierarchy that compares the results with the old system and the adaptive model by using courseware as a guide. This courseware was developed by an expert. This research can be guided by a hierarchical approach based on the concept of student achievement and ability. It can also enhance the image of the student's learning as well. Another example, a personalized mobile English vocabulary learning system that applies IRT and learning memory cycle [29]. The results from both studies suggest that IRT is a suitable tool for adaptive learning and shows potential for enhancing learning effectiveness.

From the introduction topic, the eDia system [14]. Within the system, to used technology to solve certain crucial problems in education by supporting the personalized learning has content; reading, mathematics, and science of primary education. Within the system used IRT model to establish assessment scales. The structure of the system has: Item writing, Test editing, Online test delivery, Automated scoring, Built-In Data Processing and Statistical Analyses, Teacher-Assembled Tests, Feedback and, Scaling and Setting Norms. The system tested in over 1000 schools and have item bank over 1000 items for innovation. The recommended from this research have to emphasize that an assessment instrument alone does not improve the quality of learning, it depends on how the information it provides is used to change teaching and learning processes. For better teaching and learning, there should be improvements in teaching and tools to appropriate the individual needs of the learner.

Other examples of IRT for personalized learning tend to focus on computer science courses at university. The recent study by Maddalora [10] proposes that diagnostic assessments are administered after each engagement with the source materials and IRT can calculate the “shortest learning sequence”. For the following engagement from repeated engagements, after each engagement, the materials are reduced by removing those materials that the student has already gained mastery. A personalized Web-based instruction system also developed for an introductory programming course at university by Chen [11] combined IRT with an existing courseware system. By taking into account the information value of each courseware, the system could match courseware with learner ability and thus deliver "personalized curriculum sequencing".

Wauters et al. [30] reflect upon the potential for adaptive item-based ITS based on IRT and suggest that the two challenges are: the data and the algorithm. In an ITS, responses are collecting in a less structured manner because the learner usually has some choice over which content to consume, at what pace, and when to stop. If there is insufficient coverage of responses (missing data) then it is possible that the IRT analysis cannot be performed. Wauters suggest some modifications to the ITS to ensure that this does not happen. The second challenge is designing a suitable algorithm, which can be separated into: item difficulty estimation, ability estimation, and item selection. These challenges arise because the objective of a learning environment (ITS) is different from the objective of an assessment (CAT). In an ITS, the aim is to optimize learning efficiency and maximize the learning outcome. Hence, as Wauters concludes, the algorithms used in an ITS will require modification and solutions found to the challenges.

3 RESEARCH METHODOLOGY

3.1 Preparation of the algorithm and content

For the apply the Thai language reading assessment tool for analyzing the difficulty and discrimination parameter of learning content using IRT. Propose an adaptive learning algorithm that selects questions and interventions for a learner using their predicted ability calculated by IRT. The content to practice in KidLearn was devised by experts in Thai language learning with a focus on letter sounds for children aged 6-8 years old. The content consists of three types of training games that will be implemented as the iPad application:

1. “Letter sound training” is practice in understanding the beginning sound of the word (98 questions).
2. “Word Segmenting or blending 1” is training about mixed words between consonants and vowels (252 questions).
3. “Word Ending or blending 2” is training about mixed words between consonants, vowels and spelling (56 questions).

So, the level of training game has easy (question picture like student learning pattern), medium (question picture not same as student learning pattern) and hard levels (choice similarly). The application will require a learner to login (or their parent/teacher to login). The learner will be required to take some test questions prior to receiving their personalized lessons. In this intervention we focus on “Letter Sound” training. The content consists of 98 questions, divided into 6 topics based on similar sounding letters (from 42 Thai consonants [31]). The topics are ordered by experts from easy to difficult, consisting of: topic A (21 questions), topic B (18 questions), topic C (21 questions), topic D (22 questions), topic E (9 questions) and topic F (7 questions).

The algorithm calculates the ability of a child in a particular topic according to their responses to items in the topic and the difficulty and discrimination of each item. In this calculation, the difficult and discrimination parameters are a measure of how useful the item is in differentiating between participants of high and low ability. The discrimination parameter and the difficulty parameter define the Item Response Function (1) which is represented by the Item Characteristic Curve (ICC).

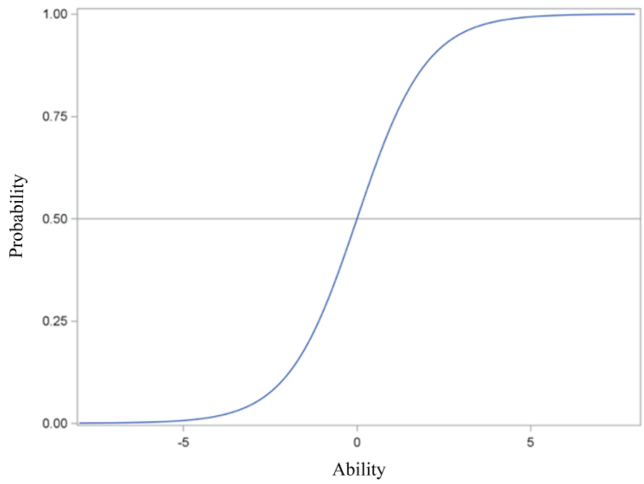


Figure 2. Item Characteristic Curve

The values of the discrimination and difficulty parameters affect the slope and the horizontal offset of the characteristic curve, respectively. A high difficulty value indicates that the item a person of higher ability is more likely to answer correctly. A high discrimination value indicates a stronger classification power. The two-parameter logistic model (2PL) calculates the probability from the difficulty and discrimination of each item [9].

$$P_j(\theta) = \frac{e^{Da_j(\theta - b_j)}}{1+e^{Da_j(\theta - b_j)}} \quad (1)$$

Where $P_j(\theta)$ is the probability that the participant will give the correct response to item j, a_j is the discrimination parameter of the item, b_j is the difficulty parameter of the item and D is a constant value of 1.702.

The discrimination parameter is a measure of how useful the item is in differentiating between participants of high and low ability. A high value indicates a high classification power. The discrimination parameter of an item can range from -∞ to ∞, but a typical value is between 0.0 and 2.0, with values closer to 0 meaning the item has a low classification power. A negative value identifies an item that has no classification power and therefore should be removed from test).

The difficulty parameter is a measure of the difficulty of the item. The difficulty parameter of an item can range from -∞ to ∞, with 0 meaning that 50% of the participants responded correctly. If the difficulty value is negative, then more than 50% of the participants responded correctly to the item. Normally, the value is between -2.8 and 2.8 [32], which can then be used to interpret the difficulty parameter as: less than -2 is very easy, -2 to -1 is easy, -1 to 1 is medium, 1 to 2 is hard and more than 2 is very hard.

In general, calculating the estimation of each child's ability uses the Maximum Likelihood Estimation (MLE) method applied with the Newton-Raphson method to calculate the probability maximum ability of the child, as in formula (2).

$$\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum_{i=1}^{N} -a_i[u_i - P_i(\hat{\theta}_s)]}{\sum_{i=0}^{N} a_i^2 \, P_i(\hat{\theta}_s) \, Q_i(\hat{\theta}_s)} \quad (2)$$

Where $\hat{\theta}_s$ is the estimated ability of the child within iteration s, a_i is the discrimination parameter of item, u_i is response for item i and N is the number of responses, $P_i(\hat{\theta}_s)$ is the probability of the correct response to item i from ICC in equation 1 at ability level $\hat{\theta}$ within iteration s. $Q_i(\hat{\theta}_s)$ is the probability of incorrect response to item i calculate by 1- $P_i(\hat{\theta}_s)$.

In CAT, the above model is applied per test (for a specific bank of questions). The proposed algorithm applies the model across multiple topics (where each has its own questions) and therefore the overall ability of the child within the system can be obtained from equation (3).

$$\bar{\theta}_s = \frac{\sum_{t=1}^{T} d_t \hat{\theta}_{st}}{T} \quad (3)$$

Where $\bar{\theta}_s$ is the averaged estimated ability of the child across T topics and $\hat{\theta}_{st}$ is the ability in topic t within iteration s. The goal of the algorithm is to maximize $\bar{\theta}_s$ for each child. At each new iteration s+1, the algorithm chooses the topic t that has the most potential to increase $\bar{\theta}_s$. The parameter d_t enables topics to be weighted independently.

For the obtaining discrimination parameter and difficulty parameter of each question calculated form the children response (In some question has missing response from children). So, we used Corrected Item Mean Substitution (CM) Imputation to estimate of item response [33]. Calculate weight function by person mean and item mean as in the formula (4).

$$\tilde{x}_{ij} = \frac{PM_i}{\frac{1}{\#obs(i)} \sum_j IM_j} IM_i \quad (4)$$

Where $\tilde{x}_{ij}$ is the estimate item response of the item j of examinee i (value is between rounded 0 or 1), PM_i is the person mean of examinee i, IM_j is the item mean of item j, $\#\text{obs}(i)$ is number of nonmissing item responses for examinee i.

In KidLearn there are 17 topics and equal weighting is applied to each topic. Therefore, the algorithm chooses the topic where the child has the lowest ability after administering each iteration of 10 items as illustrated in Figure 3. The ability in a topic is calculated from all the responses in that topic—at the end of the first iteration there will be 10 items, and the next time that topic is administered there will be 20 items, and so on the IRT model provides the information: a) to select the items that should be delivered; and b) to determine when enough responses have been gathered to confidently estimate the student’s ability. Given that a student’s ability can be obtained for the 17 topics (6 topics for Letter Sound), the information enables c) to compare abilities across topics to determine in which topic is the student weakest. This is crucial information that provides a new method to adaptively select the items and interventions that would most benefit the student.

The algorithm proposed in Figure 3 can be separated into two parts: initialization, and training. The initialization and assessment parts are common to CAT that utilize IRT, with the modification that items are grouped into topics. When the algorithm starts, there is zero information available on the child so the ability in each topic is set to zero. From here, items are repeatedly selected and delivered to the child and the ability θ_x for all *x* are recalculated. The confidence C_x in the ability for a particular topic *x* is the sum of the differences in ability between items, for the last *k* items, as shown in (3).

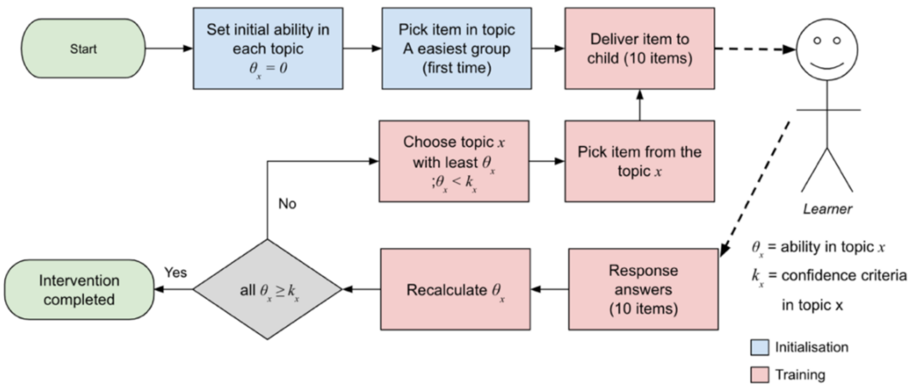


Figure 3. The algorithm design with IRT of system

A lower value implies a greater confidence, due to the differences between subsequent θ_x is relatively small.

$$C_x = \sum_{i=n-k}^{n} \left| \theta_{x,i} - \theta_{x,i-1} \right| \tag{5}$$

The first time the KidLearn application is used, there is no response data to calculate the difficulty and discrimination parameters—the so called “cold-start” problem [34]. To obtain initial parameters, three experts evaluated the 98 questions and classified each question into “easy” (question picture like student learning pattern), “medium” (question picture not same as student learning pattern) or “hard” (choice similarly) . They worked independently and reviewed their classifications together to obtain a final agreement on the classification. The experts were: Issarapa Chunsuwan (Expert in Developmental and Behavioral Pediatrics, Faculty of Medicine, Thammasat University), Kanokporn Vibulpatanavong (Lecturer in Special Education, Srinakharinwirot University), and Nichara Ruangdaraganon (Expert for Developmental and Behavioral Pediatrics, Faculty of Medicine Ramathibodi Children's Hospital). After the experts agreed the difficult classification, the items were assigned an initial difficulty value: easy is -3.0, medium is 1, hard is 3.0. The discrimination parameter was set an initial value of 0.5 for all items, which scores given at levels are based on the difficulty parameter of the IRT theory [12]. These default values were used for the initial iteration of the experiment and were then replaced with values calculated from actual responses (when each item has enough responses to calculate the parameters with IRT). In this way, the system is able to operate in the early cold-start phase.

After some iterations of selecting and delivering items (the assessment part), C_x will begin to tend towards zero for each topic *x*. At the point where all C_x are below a defined threshold t, the algorithm has performed sufficient assessment that it can begin training. The topic with the least θ_x is chosen and an intervention from the topic is recommended to the learner. An intervention could be any learning material or activity. After the intervention is completed, the algorithm returns to the assessment and delivers items to the user from the same topic as the intervention. If the learner has successfully improved then their θ_x will have increased, leading to more challenging items being delivered in future iterations of the algorithm. When the ability values settle, then the next lowest ability topic will be selected for an intervention. In this way, the algorithm "tick-tocks" between assessment and training as the learner's ability increases.

Whereas the IRT method is completed when there is sufficient confidence in the assessment estimate, the purpose of this algorithm is to increase the learner's ability across all topics and does not have a definite completion point. It can be ended or paused at any time. The algorithm is validated by considering its potential to propose suitable interventions to 10 randomly chosen samples.

3.2 Designing the KidLearn platform

The development of the KidLearn platform is divided into 3 main sections; a) backend server/database (collect bank item and algorithm), b) mobile application (training section and interaction with students) and website (display overall student result).

3.2.1 Database server side of the KidLearn platform

In the backend designed to store various important information as shown in Figure 6. The questions to training (test item bank) collect on the items table assessment has the discrimination parameter and difficulty parameter calculated by an algorithm from equation (1), which main processing part of the IRT algorithm. The scores each training and children ability collected on the response group, child ability, and child ability history respectively.

The API is also designed to serve as an intermediary that allows programs to data transfer and receive data between mobile application and web application. From the Figure 4. an example for get items from database via API, which JSON question is response request 10 items/time (for adaptive learning) from API to mobile. The model questions to map JSON are assessed to display on mobile screen.

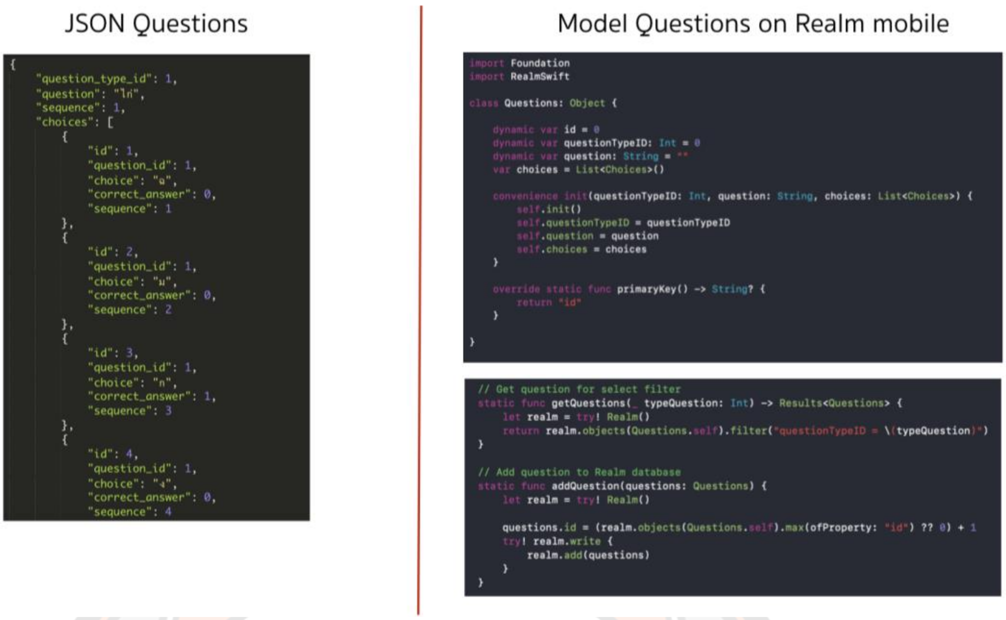


Figure 4. API request information from the Web Service.

```
func fetchChildAbility(withID id: Int, completion: @escaping(ChildAbilityResult) -> Void) {
    let route = APIRouter.childAbility(id: id)
    Alamofire.request(route).responseData(completionHandler: { response in
        print(response)
        let decoder = JSONDecoder()
        decoder.keyDecodingStrategy = .convertFromSnakeCase
        let childAbility: ChildAbilityResult = decoder.decodeResponse(from: response)
        print(childAbility)
        completion(childAbility)
    })
}

func updateChildAbility(withID id: Int, topicID: Int, questionTypeID: Int, completion: @escaping(Result<Bool>) -> Void) {
    let route = APIRouter.childAbilityUpdate(id: id, topicID: topicID, questionTypeID: questionTypeID)
    Alamofire.request(route)
        .validate(statusCode: 200...300)
        .responseJSON { response in
            switch response.result {
            case .success(_):
                completion(.success(true))
            case .failure(let error):
                completion(.failure(error))
            }
    }
}
```

Figure 5. API request child ability from the Web Service.

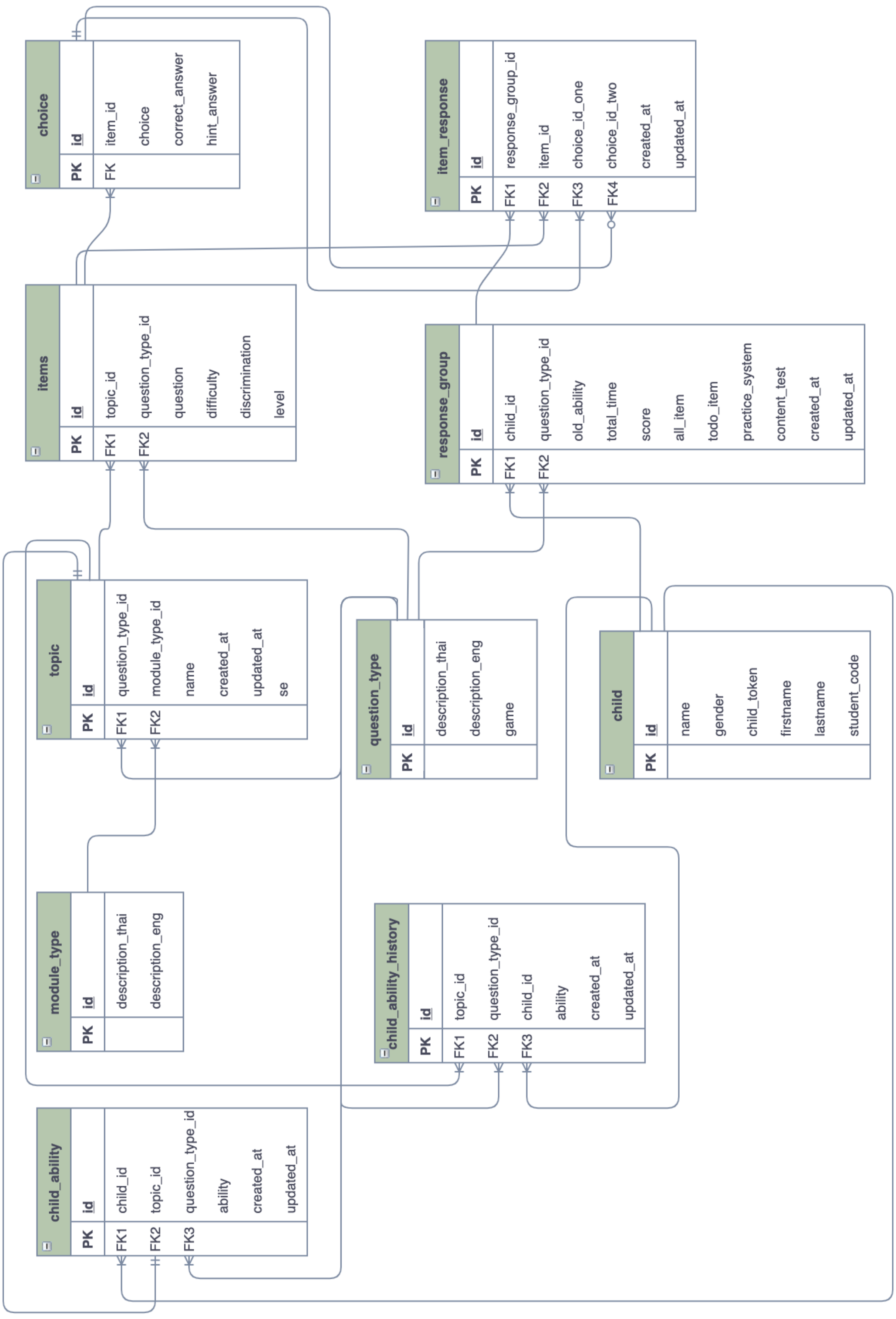


Figure 6. ER Diagram of the KidLearn database.

3.2.2 KidLearn Application

The KidLearn application design had 3 functions as shown in Figure 7; first function is learning consonants, vowels, numbers. The second function is adaptive learning (mentioned above preparation of the algorithm and content). Each question can be served in 4 different formats: “letter song”, “train drag-n-drop”, “fruit in basket” and “alphabet balloon. The first example (in Figure 8.) is a question from Topic C in the “fruit in basket” format. The child must listen to the word and drag the letter for the starting sound to the basket. The second (in Figure 9.) is from Topic A in the “letter song” format, and involves pressing on a letter instead of dragging. The application first selects the easiest topic and administers 10 questions from that topic to the child. It chooses the next topic according to the ability of the child in each topic using the proposed personalized learning algorithm [15].

Figure 7. The KidLearn main menu learning

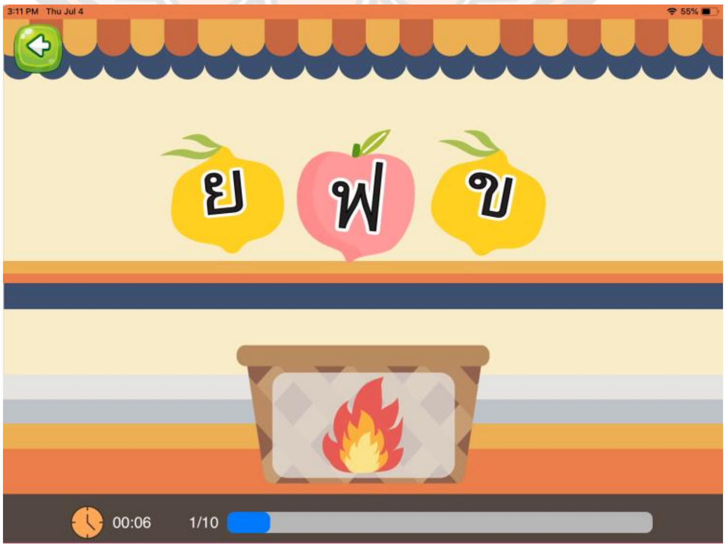


Figure 8. The fruit in basket game to learn with personalized learning

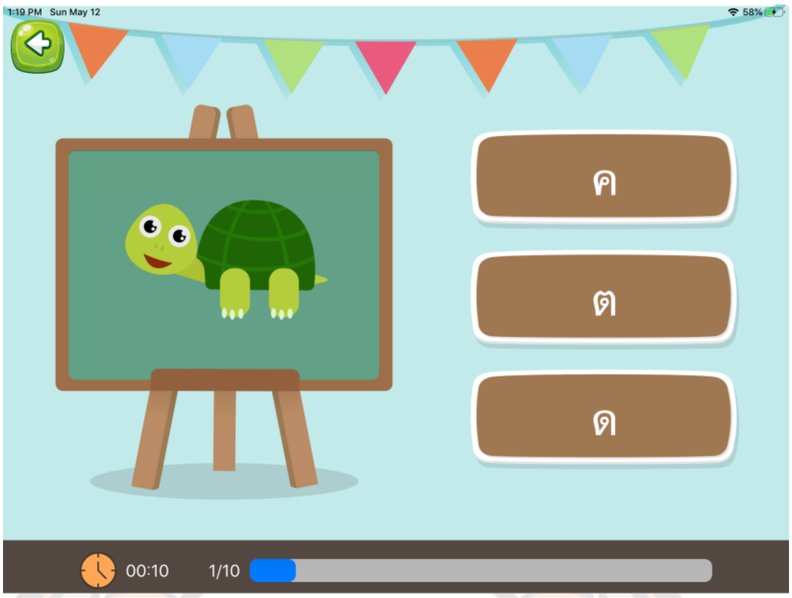


Figure 9. The letter song game to learn with personalized learning

The last function is non-adaptive learning exercises as shown in Figure 10, which lessons are divided into 6 lessons by the exports. All questions are obtained from API service and the system can collect the data when the internet disconnected. After re-connected the internet the response will be uploaded to database.

Figure 10. The review reason to learn with non-adaptive learning.

3.2.3 KidLearn website

The KidLearn website entire the display of each class of children. This website an access for teachers and experts. In order to assess the child's information and as a guideline for analysis and practice more. Which can view the test history of children by class and each person (in Figure 11.) in the child's profile page, a graph and an increase in child's abilities are displayed each time. The website is developed via PHP, and HTML language.

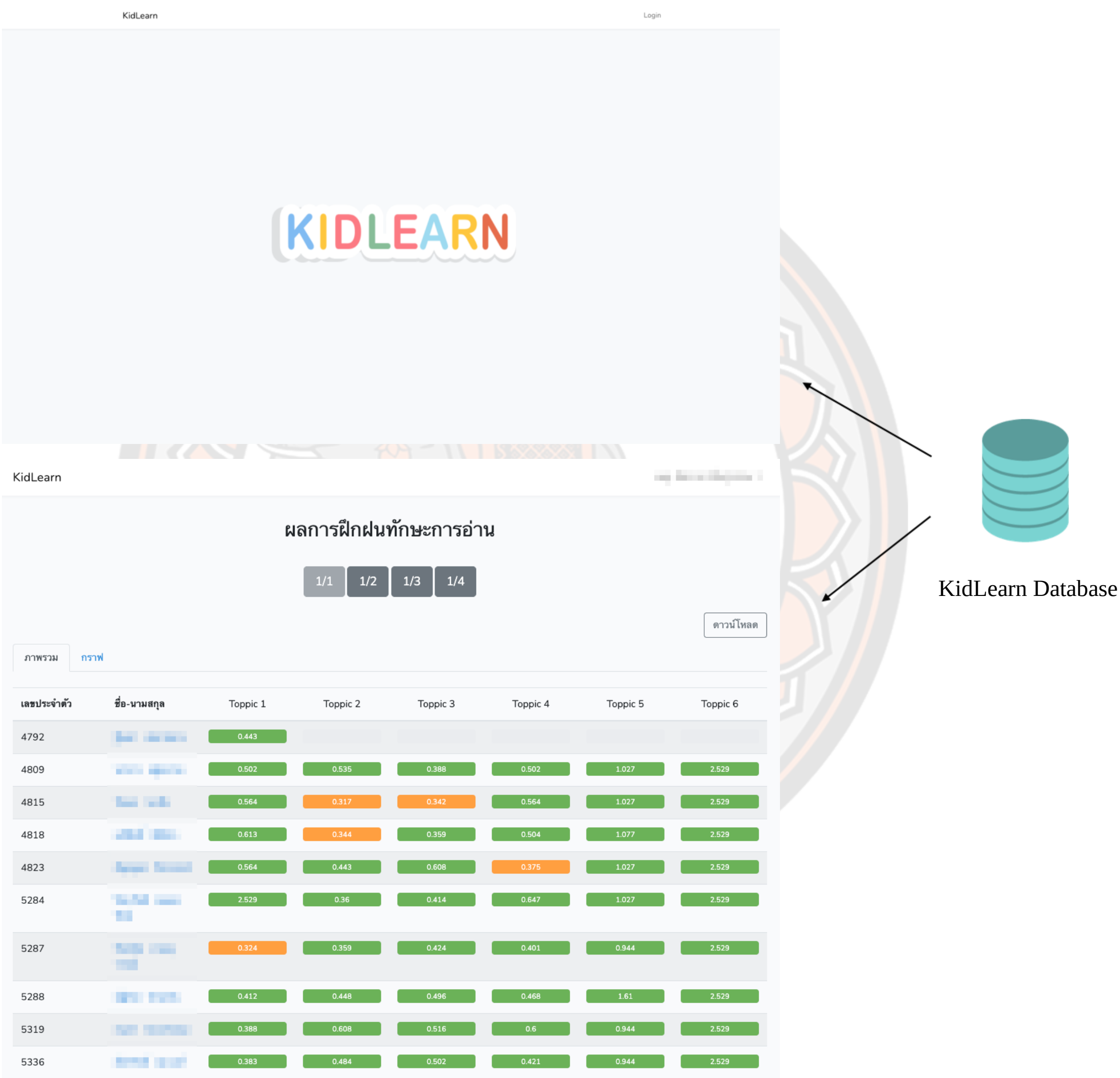


Figure 11. The KidLearn website retrieve data from database.

To view display information, the website requires login before, after logging into the admin account, under the subordinate can be displayed as Figure 12.

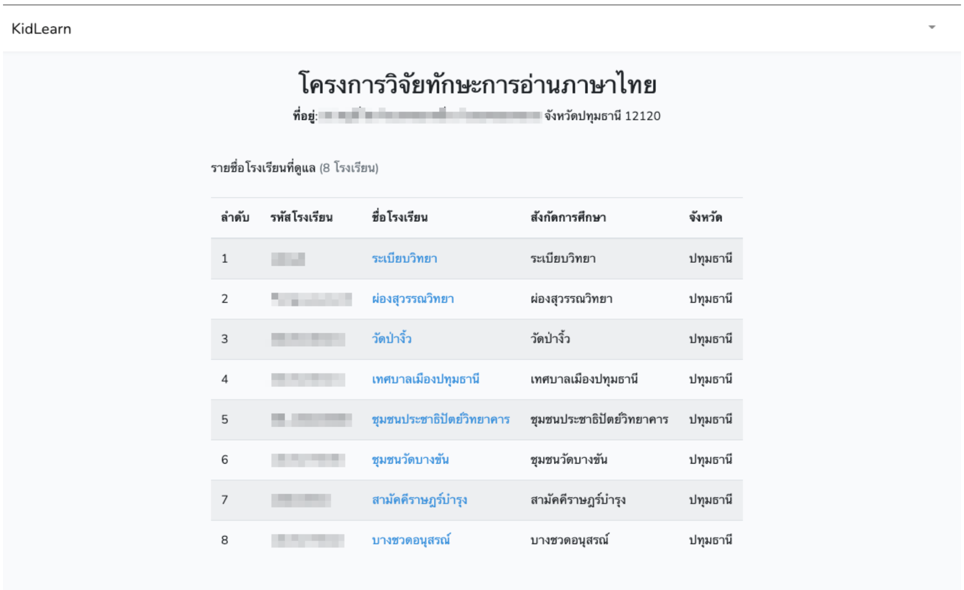


Figure 12. The page to display school group under the subordinate.

3.3 Integrating algorithm module to kidlearn platform

As mentioned above in content 3.2, we can integrate database server functionality including IRT algrithm, KidLearn application and KidLearn website as shown in Figure 13. The system will start working with the following steps:

Step 1: Learners use the Kidlearn application
Step 2: Learners login and select function 2 in Figure 7.
Step 3: The system call API get questions; which questions are derived from the calculation of the equation 5 algorithm by retrieving the data from the KidLearn database.
Step 4: After learners have response all the questions and send the data through the API,that the system will re-calculate the new learner's ability as equation 2
Step 5: End work 1 time for in-app training.
Step 6: Additional information after the test. The test results are available on the website immediately.

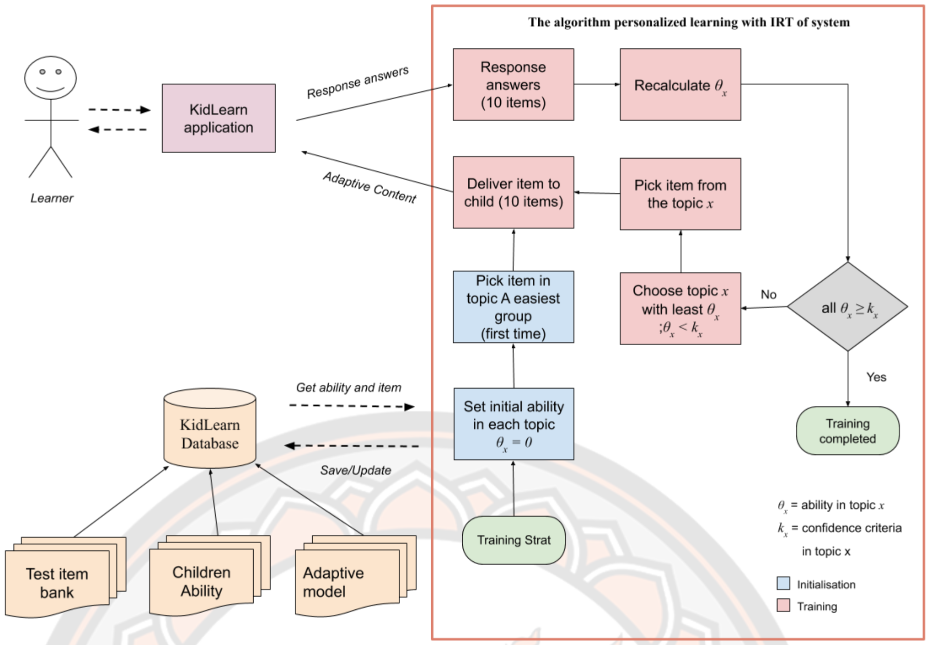


Figure 13. Architecture of KidLearn M-platform system

After each iteration of 10 items, the application sends the responses to the API and the algorithm recomputes the ability for that student. After which all the information is saved in the database on the table response group, item response, child ability, and child ability history as shown in ER Diagram of the KidLearn database as Figure 5. Then the information to display on the web applications and iPad. The teacher can monitor the response children and progress from the website. When learners train again, the system sends questions in the next section (topic) according to the child's ability to the iPad application. The system will change topic or stop when until learners pass the criteria

3.4 Experiment and data collection in schools

The KidLearn application was deployed at 3 primary schools where a preliminary assessment of the children's language ability had already been performed from KidArn application [24]. KidArn is a tablet-based application for assessing children's Thai reading skills by evaluating their ability to distinguish consonant and vowel sounds. The KidArn application was tested on normal primary student (grade 1) group to collect assessments of language abilities in students in 4 schools. The public and private schools had a total of 245 children. The results found that 47 children scored below the 10th percentile, a group considered slow learners bordering on learning difficulties. This group of 47 children was selected for the experimental group for KidLearn. The application was used by each child for 20 minutes per day (during lunch breaks) for 4 days per week (Monday to Thursday) for 4 weeks.

Each child's progress from each interaction was saved in the application and a child could come back to the same place on subsequent interactions. The algorithm always took into account all the abilities per topic $\hat{\theta}_{st}$ from previous sessions in personalizing the next iteration of questions for the interaction. KidLearn has a threshold (set by the school or experiment) for when a child can stop the activity. In this experiment, when a child's ability $\hat{\theta}_x$ reached a threshold k_x in every topic x, the sessions were no longer compulsory for the student. A threshold was set from early studies at an ability of 2.0 in every topic (beginning set ability, before calculate from children response when have enough response).

3.5 Evaluation of ability development and comparison

Two sources were integrated to evaluate the outcomes of the study: pre-post tests taken outside of the system and responses recorded within the KidLearn system. Each child in the study took a pre-test before the 4 week period, and a post-test at the end. The purpose of the pre-post tests is two-fold. First, it measures the improvement in ability of the child (at least partly) due to the intervention of KidLearn. Second, the correlation between the post test scores and the improvement in ability in each topic calculated by KidLearn gives the accuracy of the system in predicting each child's ability.

For the pre-post tests, an evaluation using Paired T-test with 95% confidence determines the significance of the child's improvement. The hypothesis is that scores on pre-test and post-test are significantly different. For the correlation, Spearman's rank correlation coefficient is calculated between the post-test scores and the ability as calculated by the KidLearn algorithm at the end of the intervention.

After the experiment, a short interview with the teacher was undertaken. The teacher was asked to share their feedback on the sessions. The following questions were asked:

1. How's child ability?
2. How child's development and training consistently?
3. How's child reading skill?
4. What's the topic problem of each child?

Do they have any problems about environment?

4 RESULT AND DISCUSSION

After the intervention, when have the data enough to calculate item discrimination parameter and difficulty parameter. We update the value the discrimination parameter and difficulty parameter of the item will make the calculation of the child's ability in each topic, which 98 questions that were collected response from 47 children. The difficulty and discrimination parameters were estimated using the R program and the “ltm” library in pre-experiment. The results are shown in Table 1

TABLE 1. THE CORRESPONDING DIFFICULTY PARAMETER AND DISCRIMINATION PARAMETER FOR EACH ITEM

Item	Topic	Level	Difficulty Parameter	Discrimination Parameter
1	A	Easy	-3.465	1.234
2	B	Easy	-3.000	1.250
3	A	Easy	-3.000	1.250
:	:	:	:	:
61	D	Medium	4.225	-0.306
62	A	Medium	-3.426	1.892
63	C	Medium	-3.485	0.779
:	:	:	:	:
96	C	Hard	-1.456	0.735
97	D	Hard	2.315	-0.572
98	E	Hard	-1.468	0.945

Table 1 (more question in an appendix) is summarize the result difficulty parameter and discrimination parameter of each item, order of item by level easy to hard (suggest from expert). The first item is topic A (easiest) has difficulty -3.465 and discrimination equal to 1.234, that this item is easy for children. The 3nd item is topic A has discrimination parameter too high 1.250. Next 98^{th} item, which is in topic E (hardest) and difficult level, but with a difficulty parameter equal to -1.468. Consequently, this item might not be the most difficult to response correctly from children answered. Typically, the estimation of the parameters of each item depends on the all correct answer of the child. The total number of answers for children who have made this item is enough that can be accurately estimated with the IRT model. The difficulty parameters and discrimination parameters of the item to change that the child's ability has changed from the default set. The system will then be able to choose items that are like children's abilities using difficulty parameters compared to children's ability.

There are changes to weak topics because some things are more difficult or some easier with the difficulty parameter of the item. The problem with data collection is the KidLearn application. The ability analysis of children using the IRT method requires the items 10-15 items/topic to be confident in estimating ability [9][35]. Due to in some topic for Thai language reading skill has less items because some topics are consonants that are rarely used. Therefore, in order to check the accuracy of the algorithm, it is recommended to repeat the tested 1 more time.

The first result is the improvement in ability as measured by the pre-post tests. Table 2 shows the improvement by topic of the 47 children that used the KidLearn system. A paired t-test indicates an improvement above 99% confidence in all topics due to a p-value of 0.003 for topic A and <0.001 for others.

TABLE 2. PAIRED SAMPLES STATISTICS SCORE PRETEST AND POSTTEST (N = 47)

Comparative issues#	Pre-Test		Post-Test		T	P
	$\bar{x}$	S.D.	$\bar{x}$	S.D.		
Topic A	8.073	2.114	9.049	1.303	3.114	.003
Topic B	5.610	1.263	6.463	0.745	5.391	<.001
Topic C	8.244	2.289	9.805	1.364	4.585	<.001
Topic D	5.951	2.224	8.171	1.548	7.040	<.001
Topic E	2.610	2.084	4.634	2.022	7.516	<.001
Topic F	1.707	1.647	3.951	2.247	8.242	<.001

The second result is that the post test results were highly correlated with the ability as calculated by the KidLearn algorithm. Table 3 shows the means of the post-test and abilities from KidLearn. The result implies that the ability determined by the algorithm is consistent with the actual ability of the child. The correlation is stronger when there are a greater number of responses, as seen from topics D, E & F which have fewer responses and hence a lower probability of correlation. This suggests that the algorithm would have more confidence in its recommendations if it set a minimum number of responses before the recommendation was enabled. As the number of topics increases this might become unfeasible and therefore some additional steps to the algorithm could flag the topics with insufficient responses to be confident of the ability.

TABLE 3. THE CHILDREN'S ABILITY BETWEEN ACTUAL ABILITY FROM POST TESTS AND ABILITY FROM KIDLEARN

	Topic A	Topic B	Topic C	Topic D	Topic E	Topic F
Post-Test mean score	8.561	6.037	9.025	7.061	3.622	2.829
Post-Test number of questions	10	7	11	11	7	7
KidLearn mean ability	19.261	15.910	18.659	18.269	7.765	5.738
KidLearn number of items	21	18	21	22	9	7
KidLearn number of responses	2317	2131	2014	1207	455	326
Correlation coefficent	0.999	0.977	0.887	0.825	0.900	0.767

The third result is that the learning sequence of topics proposed by the algorithm in KidLearn is sufficient for recommendation, but there are possibilities for improving the algorithm. In Table 4, 10 children were selected from the 47 children to show learning sequence recommended by the tutoring system from equation (5). When starting, each child has the same ability in each topic (assume a value of zero), and therefore the algorithm will select the first available topic (which is A). At the end of one intervention with items from topic A, the algorithm recalculates the child's ability for topic A: if the child performed poorly on topic A then a second intervention of topic A would follow. In most cases from Table 4, the child performed sufficiently well to obtain an ability for topic A that is above the zero level for the remaining topics, and hence at the end of topic A the algorithm selected the next available topic for the next intervention with items from topic B. Child 10 performed poorly on topic C and therefore it was repeated before moving on to topic D. After 7 interventions (covering all topics), topic C was still the weakest topic for child 10 and it was recommended twice again—the child eventually achieving sufficient ability in topic C to move onto other topics. Similarly, and highly evident, child 7 was recommended 7 consecutive rounds of topic A in order to bring their ability level on topic A up to that of the other topics. Note that if the child has similar ability in several topics then the algorithm will pick the weakest based on the IRT calculation from their responses, which may mean that they cover a wide range of topics instead of repeating one or two as is in case with child 2.

The results also show a large variation in the number of times that a child covered each topic. Child 2 completed twice as much material as child 1, despite each child being given the same amount of classroom time. Child 2 completed each round of questions faster, but with more errors, particularly in topics A and B. At the end of the classroom time, child 1 had a higher ability in topics A and B compared to child 2. The 9th child had the longest test sequence, undertaking topics A and C 8 times and 5 times respectively. Each intervention within the same topic is different, as items are randomly selected from a pool, and therefore the variation in the difficulty and discrimination of the questions that is not taken.

TABLE 4. THE LEARNING SEQUENCE OF SELECTED CHILDREN

Child#	Sequence of topics as delivered by the algorithm
1	A → B → C → D → E → F → D → D
2	A → B → C → D → E → F → B → B → B → A → A → B → B → B → B → B
3	A → B → C → D → E → F → A → D → A → D → B → B → B → B → D → A → A
4	A → B → C → D → E → F → B → C → B → B → B → B
5	A → B → C → D → E → F → A → D → B → A → A → A → A → B → B → B
6	A → B → C → D → E → F → C → B → A → A → A → A → A → A → A → C → B
7	A → B → C → D → E → F → A → A → A → A → A → A → A
8	A → B → C → D → E → F → C → C
9	A → B → C → D → E → F → A → A → A → A → A→ A → C → C → A → A → C → C → C
10	A → B → C → C → D → E → F → C → A → A → D → A → A → A → A→ A

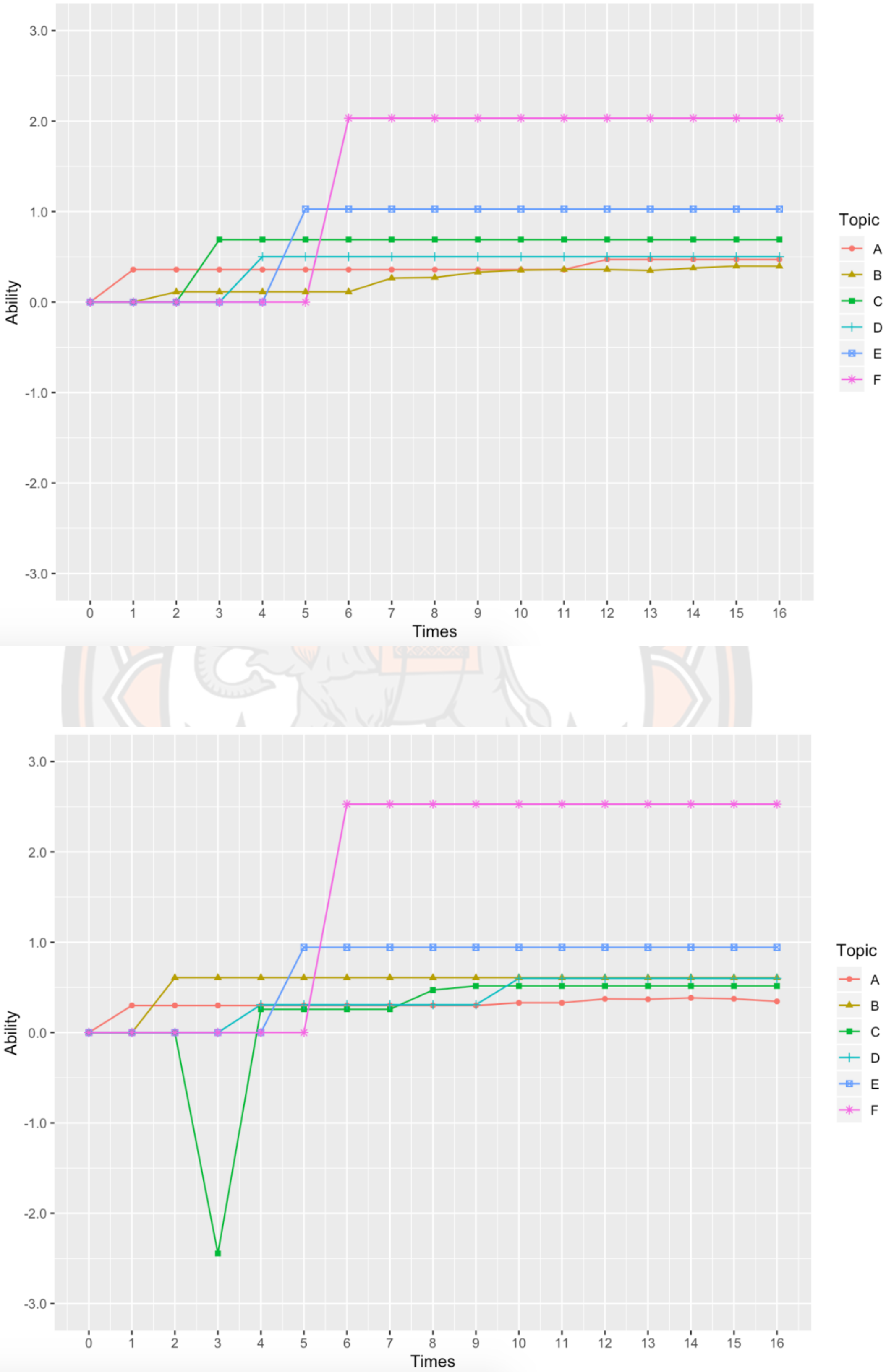


Figure 14. Ability progression during one session for child 2 (top) and child 10 (bottom)

To understand the algorithm, it is helpful to examine the abilities of individual children in each topic. From Table 4, we take Child 2 and 10 to plot their ability in each topic over time (where the x axis is interventions) as shown in Figure 14. In each intervention, the ability will change only for the topic that was recommended by the algorithm. After 1 intervention (of topic A), both perform positively, although Child 2 performs better than Child 10 (ability of topic A is ~0.4 for Child 2 versus ~0.3 for Child 10). They both also perform positively in topic B as seen by the increase in their ability. However, (as was mentioned earlier from Table 4) Child 10 performed poorly on topic C, resulting in a calculated ability for topic C of -2.5 after the first intervention. Therefore, whereas Child 2's 4th intervention was topic D, Child 10 was repeating topic C for their 4th intervention. The final topic F they both perform positively; the Child 10 performs better than Child 10 (ability of topic F is ~2.03 for Child 2 versus ~2.52 for Child 10).

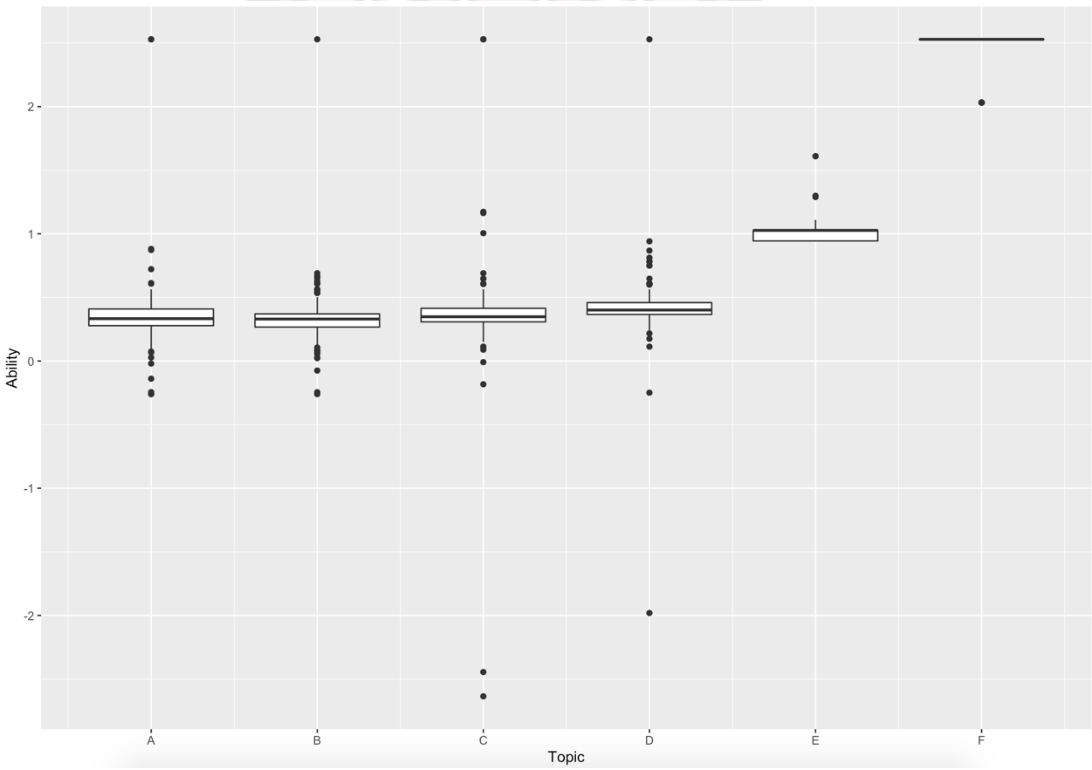


Figure 15. Ability for all children (as calculated by the algorithm) at the end of all interventions (box plot with outliers)

The abilities of all children at the end of all interventions as shown in Figure 15 indicate above average ability (average is above zero) measured against the IRT calculation performed in the pre-experiment. The average ability in topics E and F appear exceptionally high which could be explained by inaccurate or insufficient data in the pre-experiment IRT calculations, leading to sub-optimal choices of values for difficulty and discrimination for some (maybe all) of the topic E and F. An alternative explanation is that the questions in topics E & F were easier to learn for students than experts predicted. The experts chose topics A-F in terms of difficulty, with A being the easiest and F being the hardest (with the early topics being a prerequisite for later topics). Given this information, the expectation would be that Figure 15 should be an inverse relation of topic to ability. However, the data does not appear consistent with the experts' prediction of difficulty or progression. An unintended consequence of this method is that it can be used as a validation technique for experts' selection of content for topics. Further work could be undertaken to determine if the algorithm could predict which items are "out of place" in a given topic.

After the experiment, feedback was collected from teachers. Previously they had reported that it was intense work to teach spelling individually particularly blending words by themselves. After using the KidLearn app, the teachers said it was evident that the children had improved, they were reading better, some can start reading consonants, they improved their study ability and the children can back to reading at home by themselves. Therefore, the teacher's comments are correlated with the analyzes in the tables above, indicating that the lessons children have practiced with the KidLearn system can help children.

However, there was a problem that teachers reported: some children are absent school often, causing them to have discontinuous learning. When they return to practice, they received only the same topic. That is another reason why children need to practice the same topic again, as shown in Table 4.

5 CONCLUSION

In this research, the experiments undertaken on the KidLearn platform using the underlying algorithm derived from IRT with children learning difficulties (47 children). The research aims to answer that: **Can IRT be effectively applied to personalized learning of language skills for primary school children with learning difficulties?** The results can be explained as follows.

Firstly, the algorithm for selecting content based on Item Response Theory as a new technique for personalized learning is presented in Figure 3. The system can estimate the ability of children across multiple topics in a way that is consistent with pre-post test results (Table 2. and Table 3.). The results that showed the children's ability increased in each topic after using the KidLearn system could be a function of the quality of the content. However, the key result is that the ability calculated by the system is highly correlated with the actual ability of the children as determined by the pre-post test. The sequence of recommended topics could benefit from some additional rules or logic. As was evident from the results, some children experienced severe repetition of topics when they were unable to achieve an ability score above the other topics. The algorithm could be modified to avoid this repetition by not selecting any topic that is repeated x times.

In the results by topic, Topics E and F produced too high ability score which meant that they were typically only delivered for a single iteration. This was mostly caused by the cold-start approach which involved experts rating the items, and the items being easier than the experts predicted. As it turns out, the IRT approach is well-suited to detecting errors in the expert's predictions. However, in the case of this experiment the cold start values chosen by experts caused irreparable damage to the children's ability scores from which the algorithm could not recover. In a future experiment it would be important to understand how to choose the initial parameters for difficulty and discrimination when the system does not have sufficient responses to calculate itself.

After the experiment, we re-calculated difficulty and discrimination parameters found that was some error in the results (Table 1.) that difference with set initial value from the experts. Due to, we chosen as the group for the experiment. A more realistic set of difficult and discrimination parameters would be calculated from a sample of responses from the population of students instead of only learning difficulties students. If the system was deployed to the entire school (or an entire district) then the ability scores for each child would make a meaningful comparison to determine when children's learning difficulties were overcome.

Secondly, the KidLearn mobile application and backend system for training students with learning difficulties in Thai language as an example of IRT for personalized learning. The data analysis results and feedback of the KidLearn platform show potential for further research into delivering personalized learning that is based on a mathematical approach to recommending content—either in the refinement of the algorithm or in delivering different domains of learning material via platform.

Overall, from the all result mentioned all above, conclude the IRT can be effectively applied to personalized learning of language skills for primary school children with learning difficulties and can improving to more effective. In the future, the KidLearn platform hopes this research will lead to the problem solving of reading skills or other topics further.

APPENDIX

TABLE 5. THE MODULE TYPE TABLE FOR NON-ADAPTIVE TRAINING

id	description_eng	description_thai
1	Modele 1	Thai Alphabet
2	Module 2	Blending Alphabet + vowels (long tone)
3	Module 3	Blending Alphabet + vowels (short tone)
4	Module 4	Blending Alphabet + single vowels
5	Module 5	Blending Alphabet + blending vowels
6	Module 6	Blending Alphabet + excess vowels
10	Module 10	Blending Alphabet + vowels + word ending (Mae Kong, Mae Kon, Mae Kom, Mae Koei, Mae Kow)
11	Module 11	Blending Alphabet + vowels + word ending (Mae Kok, Mae Kob, and Mae Kod)
13	Module 13	Word blending has vowels reduce form
14	Module 14	Word blending has vowels change form

TABLE 6. THE ALL TOPIC FOR ADAPTIVE TRAINING (3 PART; LETTER SOUND, WORD SEGMENTING, WORD ENDING)

id	name	se	question_type_id
1	Topic A	0.36	2
2	Topic B	0.35	2
3	Topic C	0.35	2
4	Topic D	0.40	2
5	Topic E	0.94	2
6	Topic F	2.00	2
7	Vowels (long tone)	2.00	3
8	Vowels (short tone)	2.00	3
9	Vowels blending	2.00	3
10	Mae Kok	2.00	4
11	Mae Kob	2.00	4
12	Mae Kod	2.00	4
13	Mae Kod	2.00	4
14	Mae Kom	2.00	4
15	Mae Kong	2.00	4
16	Mae Koei	2.00	4
17	Mae Kow	2.00	4

TABLE 7. THE ALL ITEMS FOR TRAINING BY TOPICS

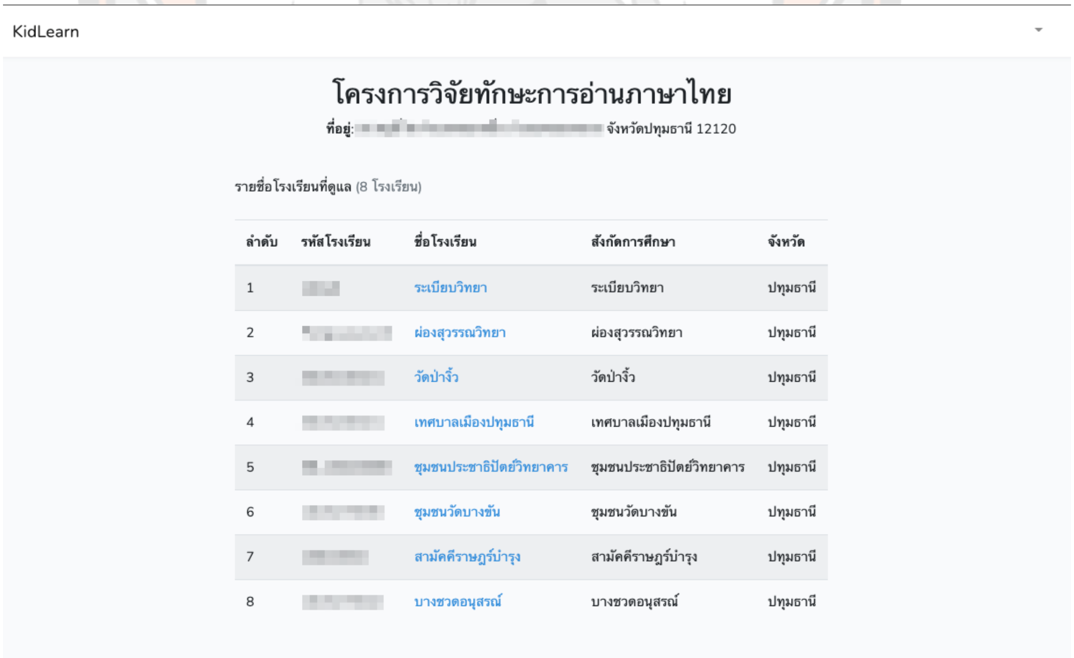
id	topic_id	question_type_id	question	difficulty	discrimination	level
1	A	2	เด็ก	-3.465	1.234	easy
2	B	2	หนู	-3.000	1.250	easy
3	A	2	ไก่	-3.000	1.250	easy
4	D	2	แหวน	-3.000	0.836	easy
5	A	2	จาน	-2.294	0.177	easy
6	C	2	ยักษ์	-2.161	0.777	easy
7	A	2	เต่า	-1.366	0.018	easy
8	C	2	ฮูก	3.877	-0.632	easy
9	D	2	ไข่	-3.356	1.001	easy
10	A	2	ใบ	-3.209	0.475	easy
11	B	2	ทหาร	4.288	-0.608	easy
12	B	2	งู	-3.000	0.709	easy
13	A	2	อ่าง	2.552	-0.956	easy
14	C	2	เรือ	-1.338	0.181	easy
15	D	2	ฉิ่ง	3.000	-0.320	easy
16	C	2	ม้า	-3.223	1.629	easy
17	B	2	ช้าง	3.281	-0.767	easy
18	C	2	ลิง	-3.453	0.745	easy
19	D	2	หีบ	3.054	-0.802	easy
20	D	2	ถุง	3.265	-0.113	easy
21	A	2	ปลา	-3.669	0.662	easy
22	C	2	พาน	4.082	-0.284	easy
23	D	2	เสือ	2.654	-0.530	easy
24	B	2	ควาย	-3.632	0.995	easy
25	F	2	หญิง	1.388	-1.241	easy
26	B	2	โซ่	2.943	-1.350	easy
27	D	2	ผึ้ง	-3.251	0.566	easy
28	F	2	เณร	1.165	-2.165	easy
29	C	2	ฟัน	-4.281	0.727	easy
30	E	2	ธง	-1.597	2.026	easy
31	E	2	ศาลา	-2.690	1.380	easy
32	E	2	ฤาษี	-2.413	1.567	easy
33	E	2	สำเภา	1.149	-2.247	easy
34	F	2	ฐาน	1.272	-1.427	easy

35	D	2	ฝา	2.349	-1.567	easy
36	E	2	ระฆัง	-2.138	1.443	easy
37	E	2	จุฬา	-3.000	0.365	easy
38	F	2	ผู้เฒ่า	1.803	-0.901	easy
39	E	2	มณโฑ	2.513	-0.734	easy
40	F	2	ปฏัก	2.762	-0.414	easy
41	F	2	ชฎา	2.505	-0.952	easy
42	F	2	เฌอ	1.970	-0.572	easy
43	A	2	ดาว	1.750	1.250	medium
44	B	2	นม	-3.000	1.250	medium
45	A	2	กุ้ง	-1.392	1.250	medium
46	C	2	ว่าว	-4.209	0.686	medium
47	A	2	ใจ	1.585	-2.105	medium
48	C	2	ยา	2.000	1.250	medium
49	A	2	ตา	2.000	1.250	medium
50	D	2	ข้าว	1.230	-3.764	medium
51	A	2	บ้าน	1.368	-0.226	medium
52	B	2	ทอง	-3.353	1.118	medium
53	B	2	เงิน	2.595	-1.327	medium
54	A	2	โอ่ง	3.456	-0.690	medium
55	C	2	รถ	-3.620	0.681	medium
56	D	2	ฉาบ	-3.000	0.604	medium
57	C	2	มด	1.478	0.285	medium
58	B	2	ชาม	3.526	-0.859	medium
59	C	2	ล้อ	2.000	1.250	medium
60	D	2	หอย	1.727	-0.869	medium
61	D	2	ถ้วย	4.225	-0.306	medium
62	A	2	เป็ด	-3.426	1.892	medium
63	C	2	พัด	-3.485	0.779	medium
64	D	2	เสื้อ	-4.721	0.426	medium
65	B	2	ค้อน	2.158	-0.906	medium
66	B	2	ซิป	4.294	-0.664	medium
67	D	2	ผี	-3.004	0.813	medium
68	C	2	ไฟ	2.114	-1.511	medium
69	D	2	ฝน	1.563	-1.083	medium
70	E	2	ฆ้อง	-1.056	2.179	medium
71	A	2	ดิน	2.872	-0.658	hard
72	B	2	นก	4.078	-0.516	hard

73	A	2	กบ	2.217	-1.232	hard
74	C	2	แว่น	1.609	-2.909	hard
75	A	2	จอบ	2.866	-1.412	hard
76	C	2	ยุง	3.000	-1.250	hard
77	A	2	โต๊ะ	3.000	-1.250	hard
78	D	2	เข็ม	1.067	-0.632	hard
79	A	2	บัว	0.730	-0.583	hard
80	B	2	ทาง	3.000	1.250	hard
81	B	2	เงาะ	3.000	1.250	hard
82	A	2	อ้อย	2.195	-1.182	hard
83	C	2	รุ้ง	2.931	-1.067	hard
84	D	2	ฉีด	0.411	-1.276	hard
85	C	2	แมว	3.000	1.250	hard
86	B	2	ช้อน	-4.047	0.418	hard
87	C	2	ลา	5.161	-0.324	hard
88	D	2	เห็ด	-2.829	0.759	hard
89	D	2	ถั่ว	0.356	-1.960	hard
90	A	2	ปู	-3.291	0.462	hard
91	C	2	แพะ	-4.406	0.640	hard
92	D	2	ส้ม	3.470	-0.738	hard
93	B	2	ค้อน	3.000	1.000	hard
94	B	2	ซิป	-4.267	0.598	hard
95	D	2	ผม	-2.429	0.298	hard
96	C	2	ฟาง	-1.456	0.735	hard
97	D	2	แฝด	2.315	-0.572	hard
98	E	2	ฆ้อง	-1.468	0.945	hard

KidLearn website

Figure 16. The KidLearn home page



ลำดับ	รหัสโรงเรียน	ชื่อโรงเรียน	สังกัดการศึกษา	จังหวัด
1	[illegible]	ระเบียบวิทยา	ระเบียบวิทยา	ปทุมธานี
2	[illegible]	ผ่องสุวรรณวิทยา	ผ่องสุวรรณวิทยา	ปทุมธานี
3	[illegible]	วัดป่างิ้ว	วัดป่างิ้ว	ปทุมธานี
4	[illegible]	เทศบาลเมืองปทุมธานี	เทศบาลเมืองปทุมธานี	ปทุมธานี
5	[illegible]	ชุมชนประชาธิปัตย์วิทยาคาร	ชุมชนประชาธิปัตย์วิทยาคาร	ปทุมธานี
6	[illegible]	ชุมชนวัดบางขัน	ชุมชนวัดบางขัน	ปทุมธานี
7	[illegible]	สามัคคีราษฎร์บำรุง	สามัคคีราษฎร์บำรุง	ปทุมธานี
8	[illegible]	บางชวดอนุสรณ์	บางชวดอนุสรณ์	ปทุมธานี

Figure 17. The page to display school group under the subordinate.

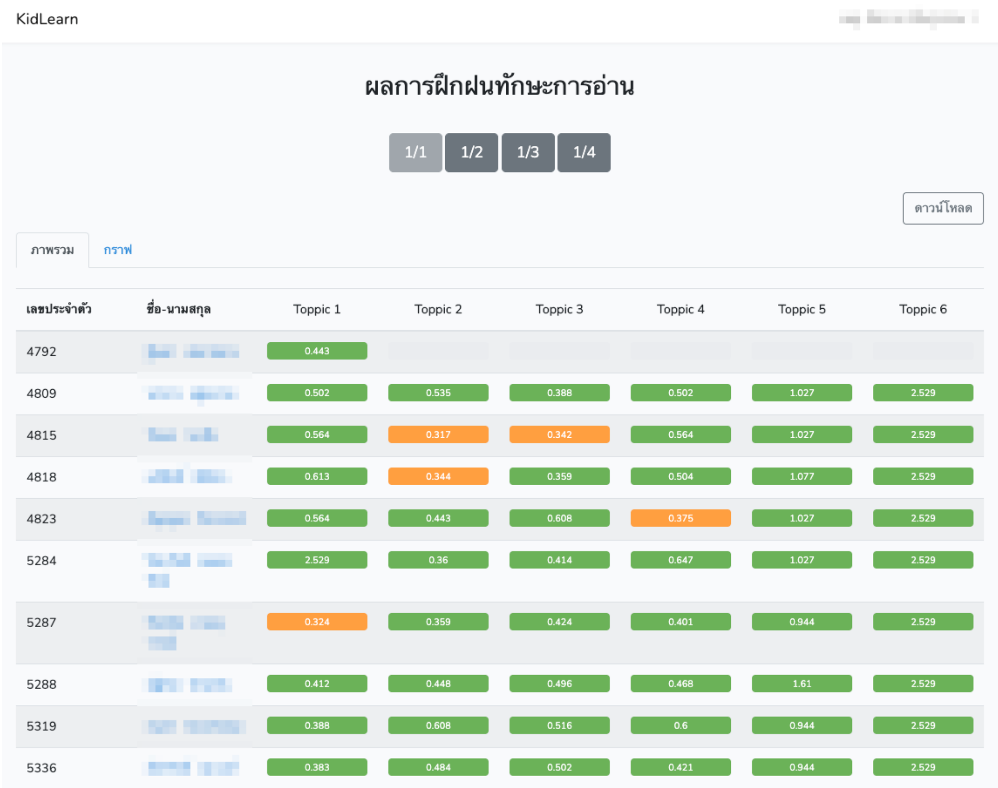


Figure 18. The student list in the school

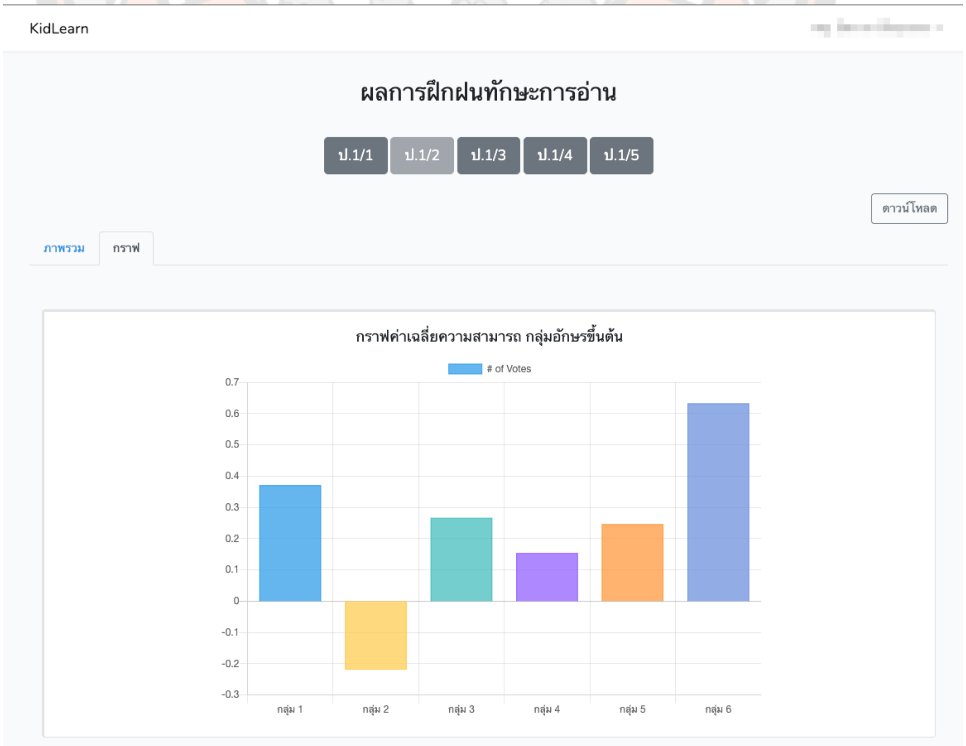


Figure 19. The graph average children ability

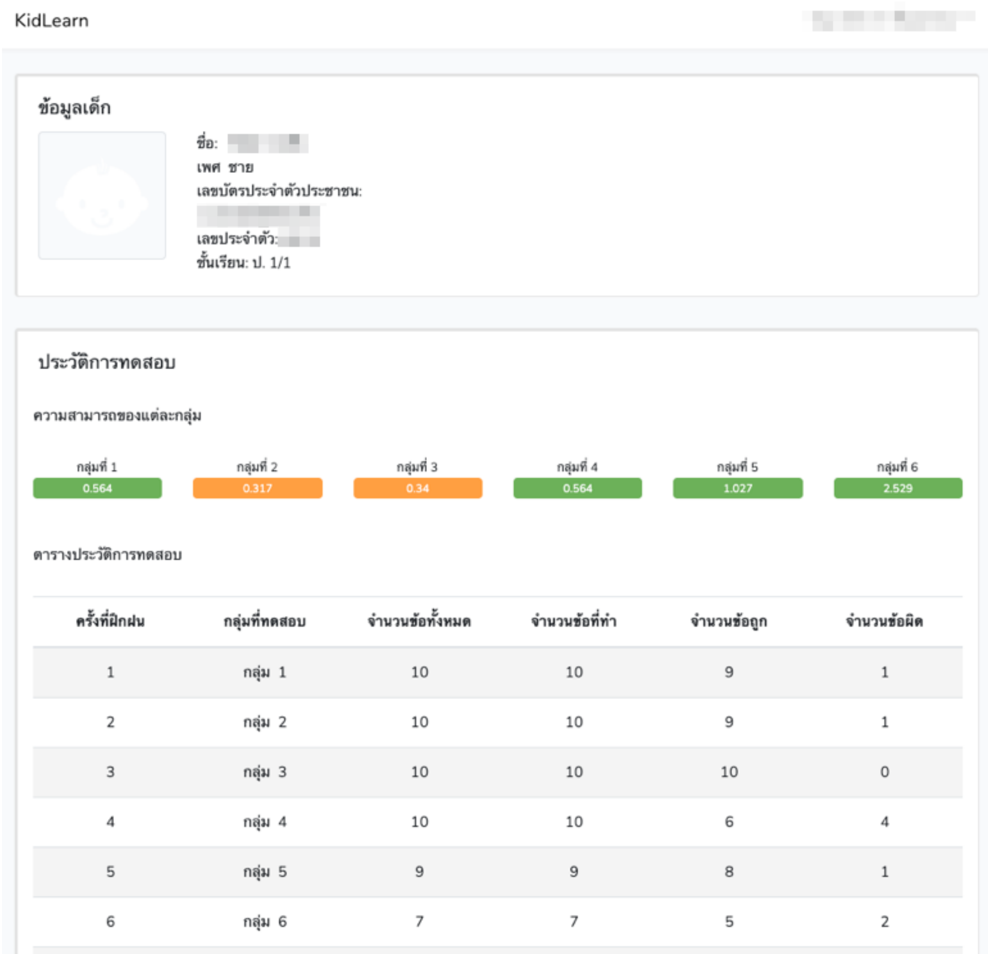


Figure 20. The child test history detail

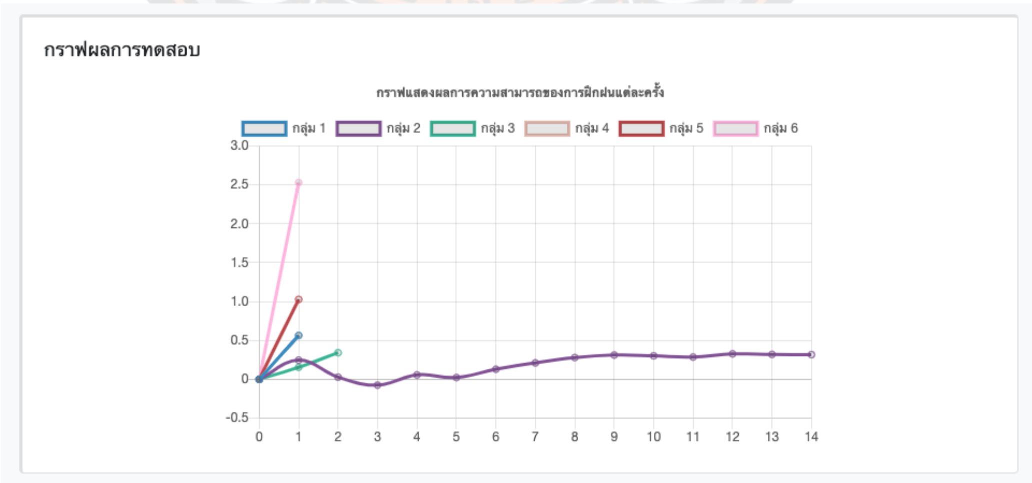


Figure 21. The learning sequence topics by adaptive learning graph

How to use KidLearn application

1. Login

To access the system, users must enter their username and password to access the KidLearn application.

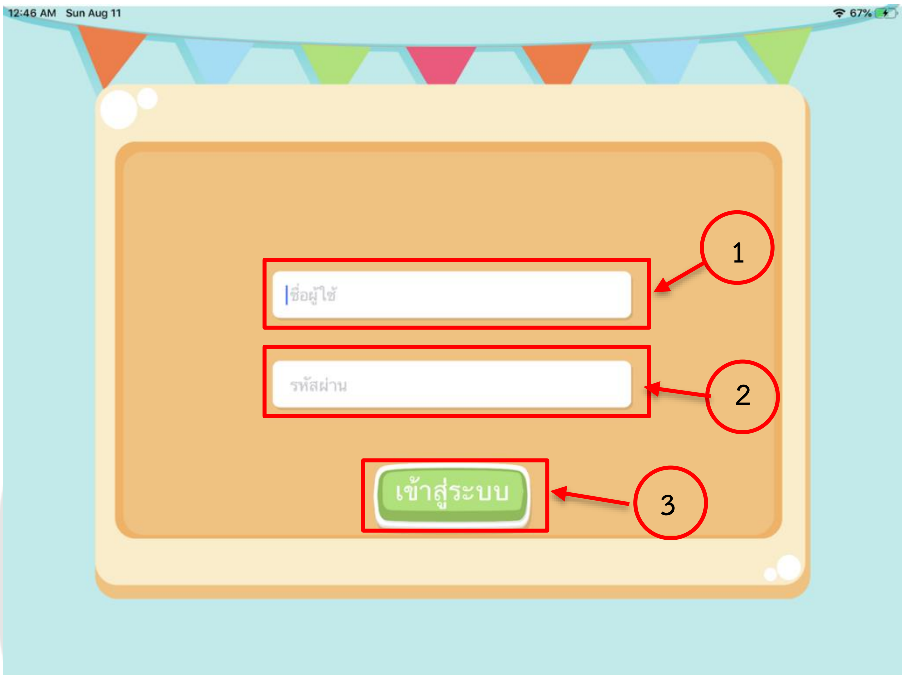


Figure 22. Login page

1. Enter your username or email
2. Enter your password
3. Press login button

2. Select training list

2.1. In the case of a school user, skip to step 2.2 if it is a school administrator Select the school first (Figure 23) and then select the test name list 2.2 (Figure 24 and Figure 25 respectively).

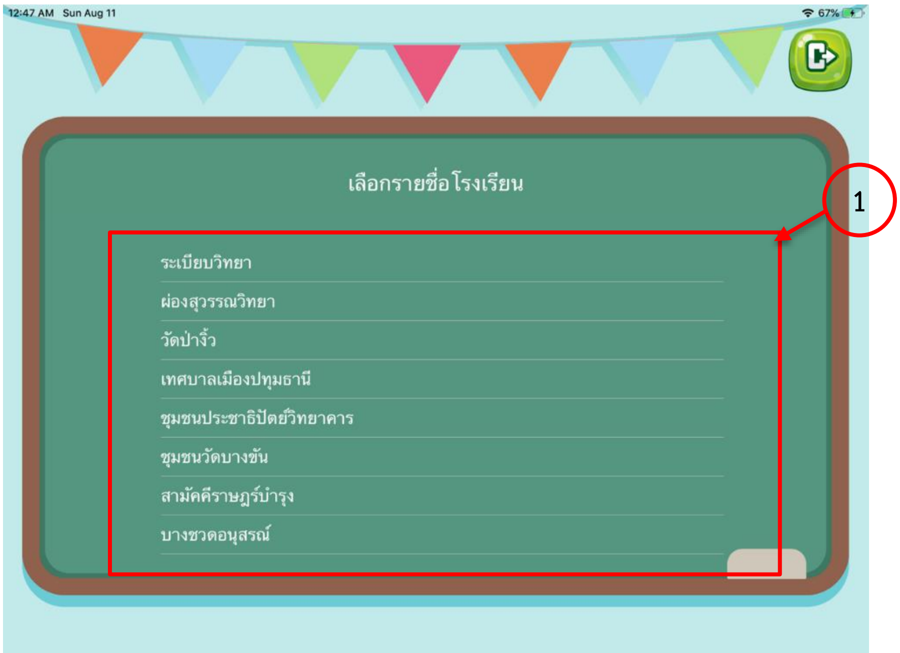


Figure 23. The page shows the list of affiliated schools.

1. Choose a school that wants to test

2.2. Select a training list

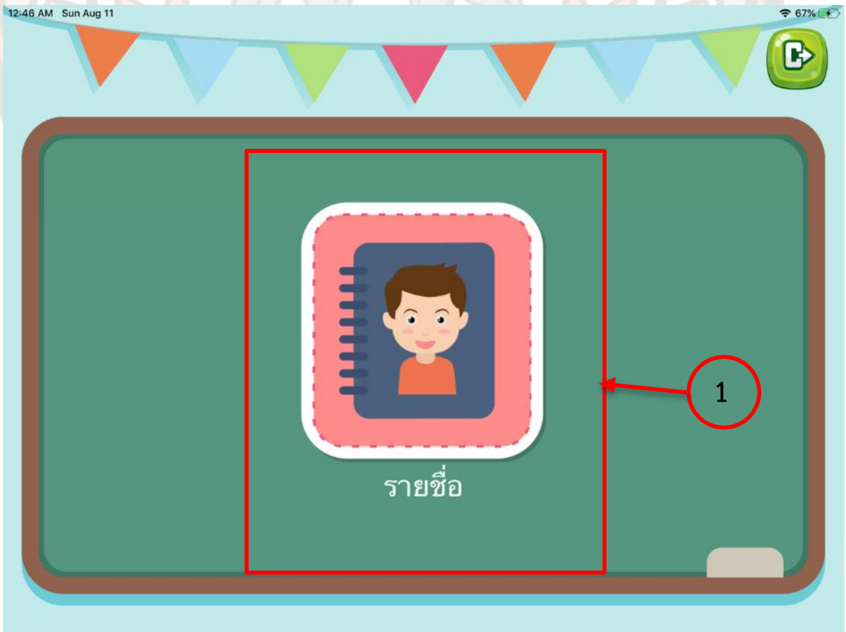


Figure 24. Search student training

1. Press list button

Figure 25. Classroom list and student names list page

2. Select the classroom to test.
3. List of students of the selected class from item 1
4. Press the practice button to train the desired student.

3. KidLearn Menu

The KidLearn app is divided into 3 main functions: learning Alphabet, training game (adaptive learning), Thai language review (non-adaptive learning). Which users can choose to play as desired.

Figure 26. KidLearn main menu

1. learning lessons button
2. Training game button
3. Review lessons button
4. Back to list button
5. Ability of children button
6. Name description

4. Learning lessons menu

The learning lessons divided into 6 groups: Alphabet ก -ฮ , Consonants, numbers, low, middle and upper characters.

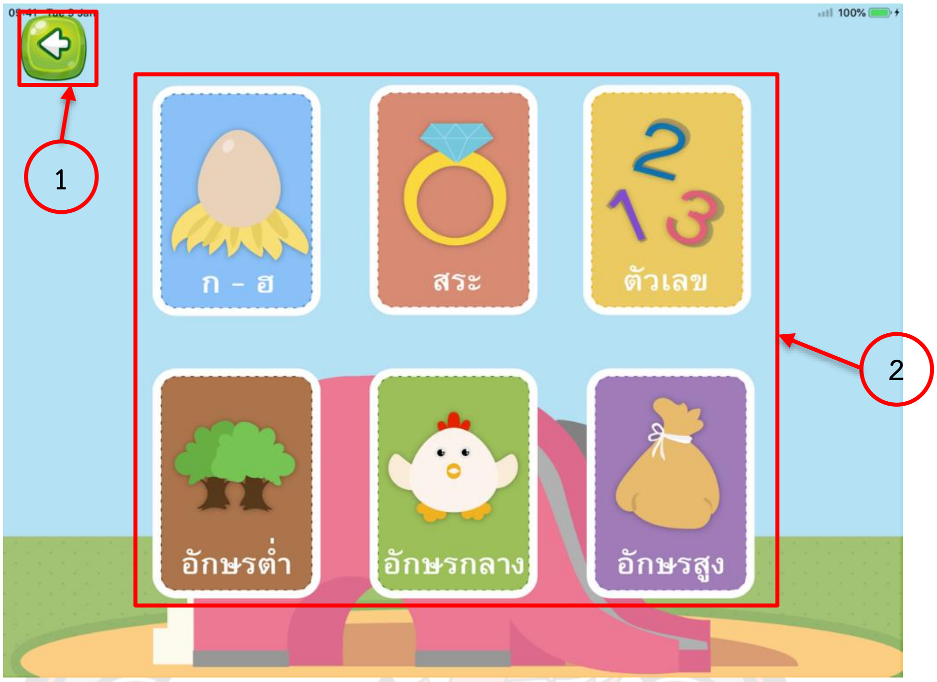


Figure 27. Learning lessons page

1. Back to main menu button
2. Select learning button

Figure 28. Alphabet learning lesson page

3. Button to select the alphabet that you want to learn more about.

4. Picture of the selected alphabet, and can press on the picture to hear the sound of the alphabet

5. Close learning lesson button

5. Training game (adaptive learning)

The training game is practice Thai language skill, that provide assessment adaptive by the children ability skill. The assessment will be start from "Letter Lound", "Initial Sound", "Word Blending or Word Segmenting", and "Word Ending", respectively.

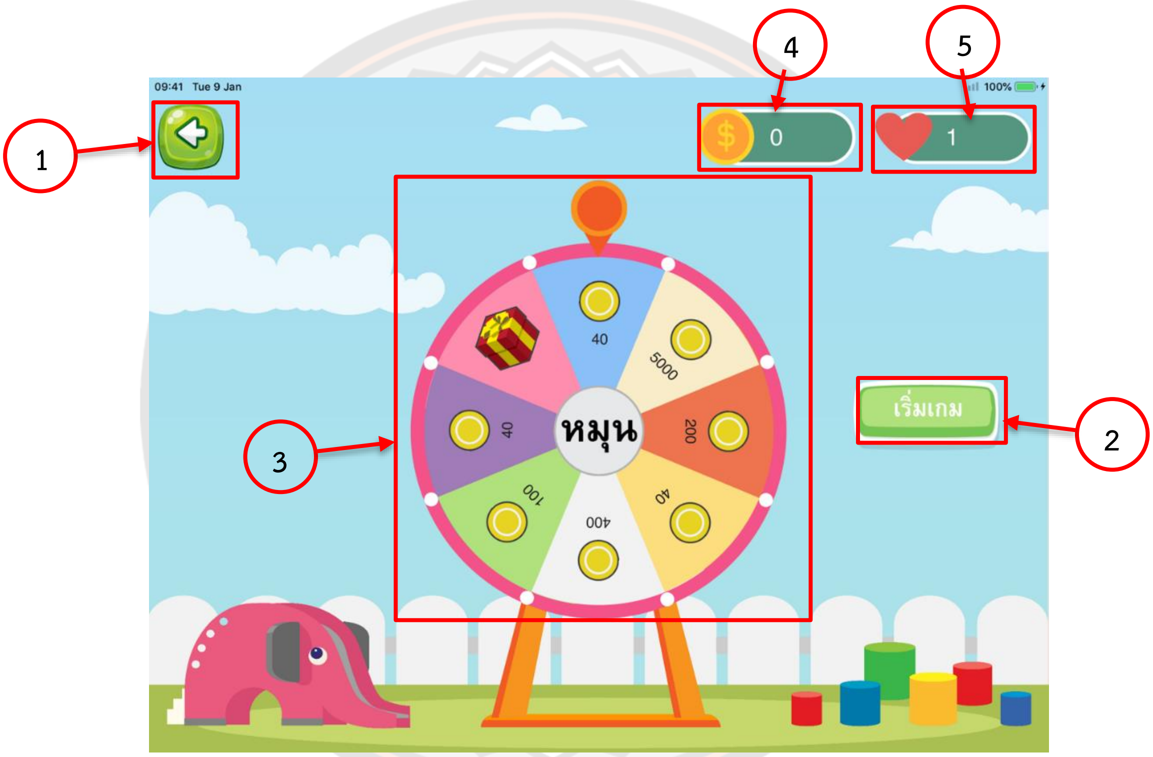


Figure 29. Main training game page

1. Back to main menu button
2. Start to training button
3. Wheel game
4. Coin
5. Heart count

6. Thai language review (non-adaptive learning)

All games are divided into 2 type: letter sound 4 games and blending 3 games

6.1. Example how to training

How to play the game explains how to play each game how to play. It will be an animation, and accompanying sound

Figure 30. Example how to training

1. Back to training menu button
2. The animation to demonstrate how to play
3. Start to training button

6.2. Letter Sound game

1. Letter Sound game 1 (Balloon)

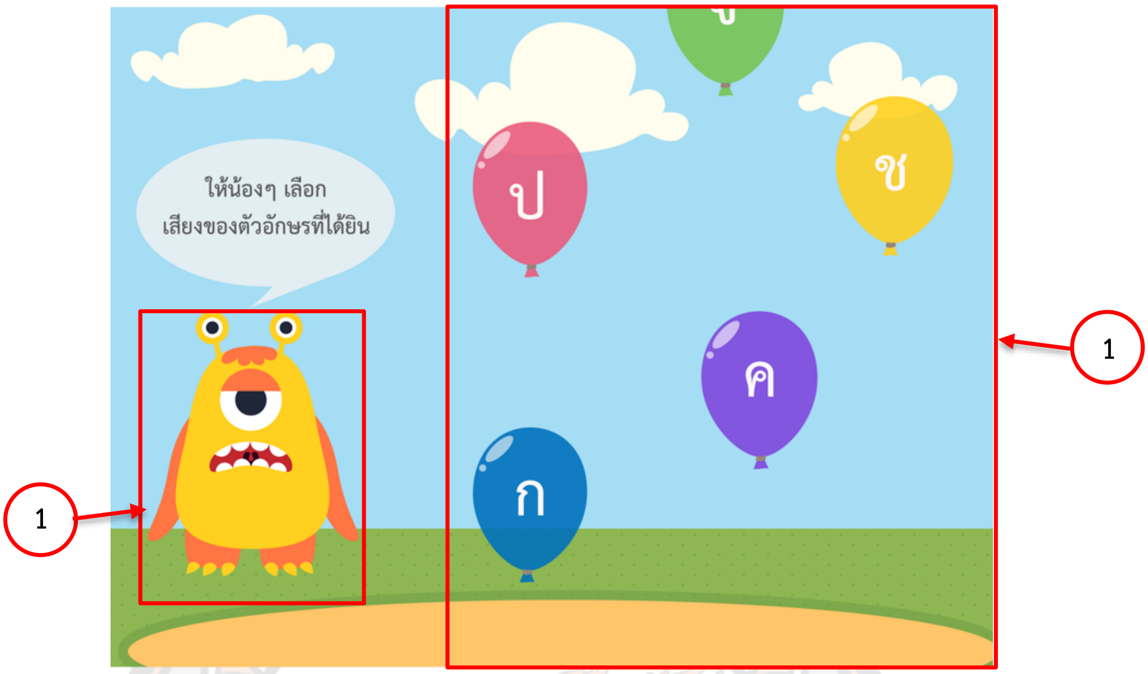


Figure 31. Letter Sound game 1 (Balloon)

1. The game explains word before start "In this game, to have children to listen the words and choose the correct consonant from the floating balloon"
2. Listen alphabet again

6.2.1. Letter Sound game 2 (Letter Song)

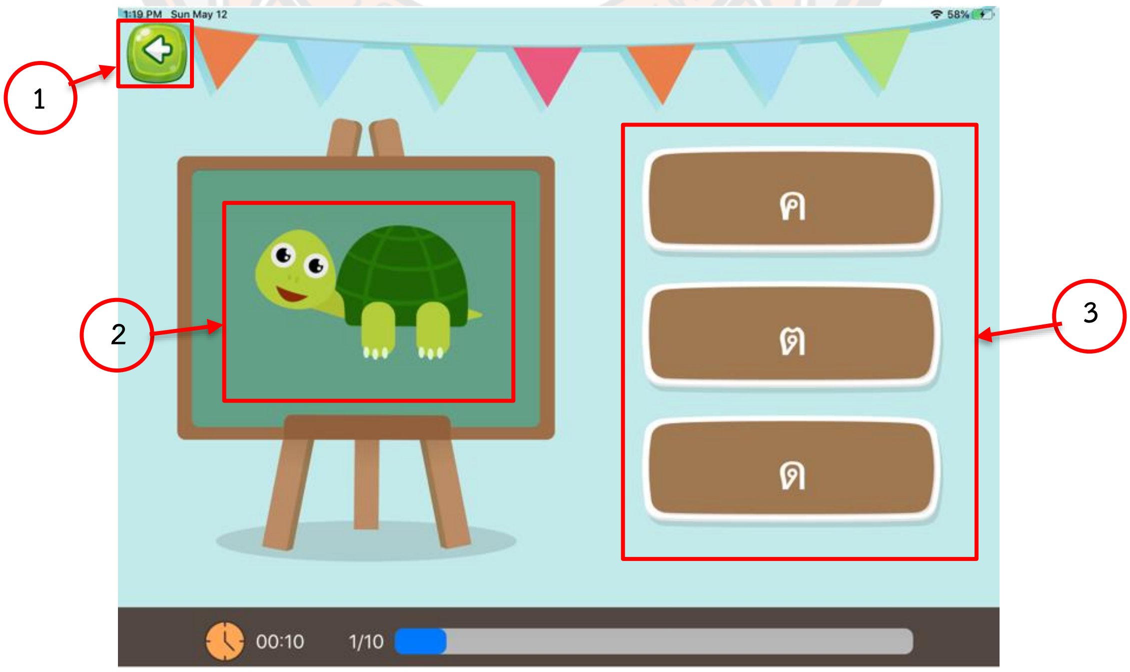


Figure 32. Letter Sound game 2 (Letter Song)

1. Back to training menu button
2. Listen alphabet again
3. The game explains word before start "In this game, to have children look at the picture and listen carefully to the name of the picture. Then choose the correct starting consonant of the picture by pressing the answer selection 3"

6.2.3. Letter Sound game 3 (Basket Fruits)

Figure 33. Letter Sound game 3 (Basket Fruits)

1. Back to training menu button
2. Listen alphabet again
3. The game explains word before start "In this game, to give children look at the picture and listen carefully to the name of the picture. Then drag the correct answer and drop to the basket"

6.2.4. Letter Sound game 4 (Train)

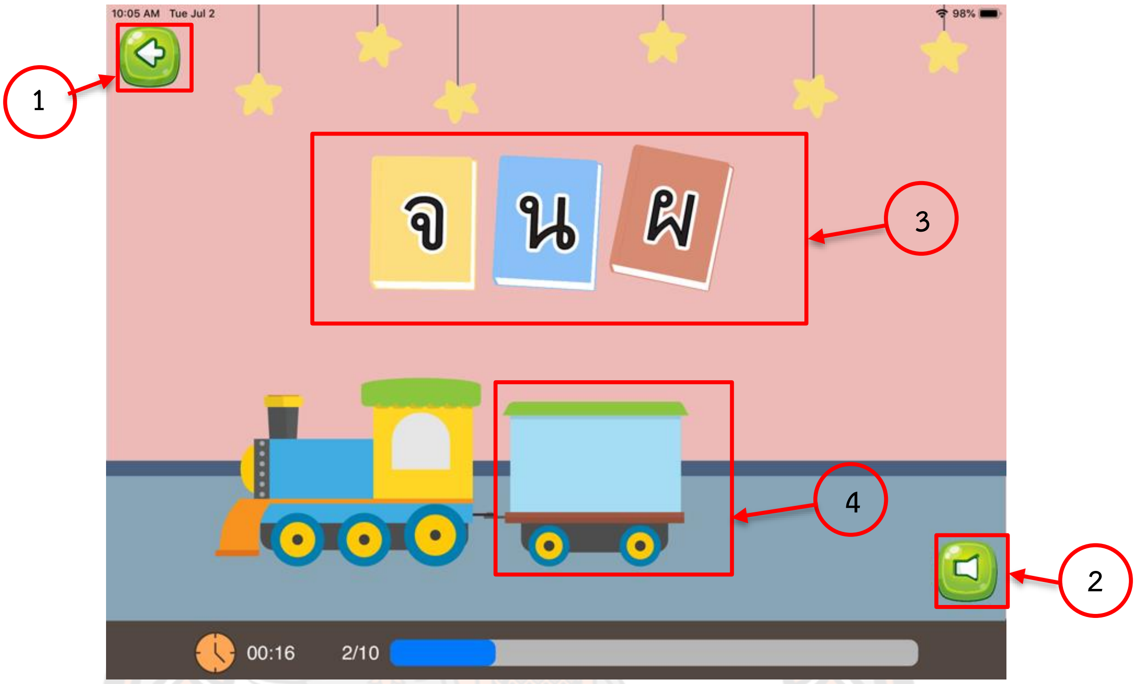


Figure 34. Letter Sound game 4 (Train)

1. Back to training menu button
2. Listen alphabet again
3. The game explains word before start "In this game, to give children look at the picture and listen carefully to the name of the picture. Then drag the correct answer and drop to the section 4"

6.3. Word Blending

6.3.1. Word Blending 1 (Shelf)

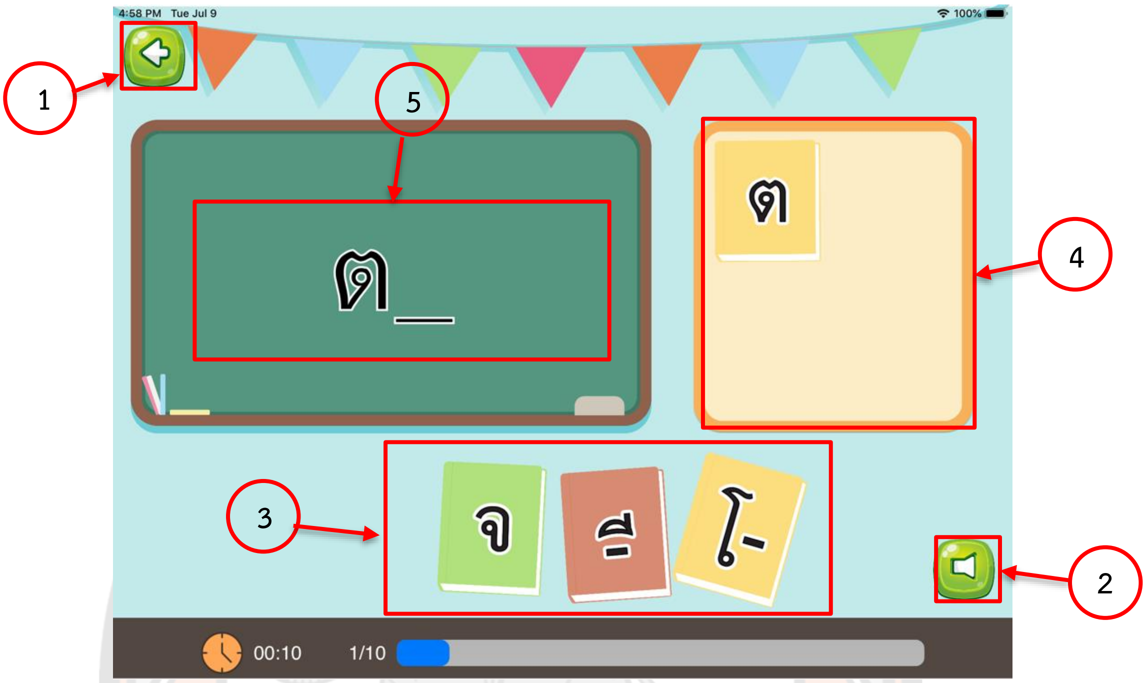


Figure 35. Screen of Word Blending 1 (Shelf)

1. Back to training menu button
2. Listen wording again
3. The game explains word before start "In this game, to give children look at the picture and listen carefully to the name of the picture. Then drag the correct answer and drop to the section 4"
4. Alphabet choices
5. Display the result word blending from the children

6.3.2. Wording Blending 2 (Basket Fruit)

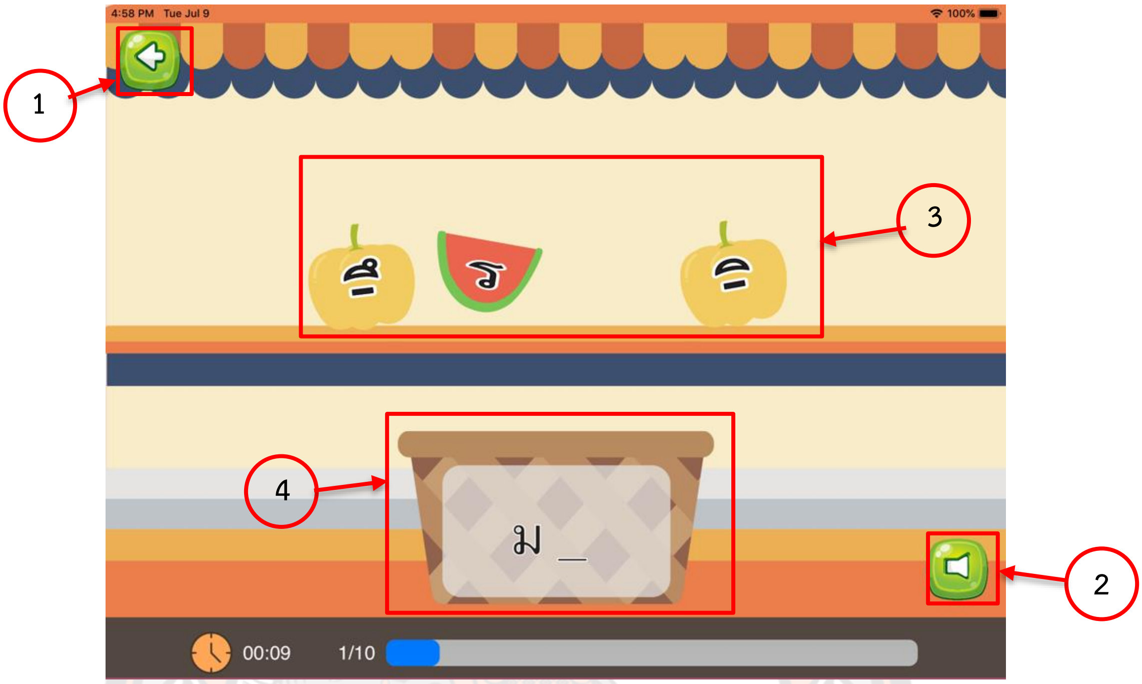


Figure 36. Screen of Wording Blending 2 (Basket Fruit)

1. Back to training menu button
2. Listen wording again
3. The game explains word before start "In this game, to give children look at the picture and listen carefully to the name of the picture. Then drag the correct answer and drop to the section 4"
4. Display the result word blending from the children

6.3.3. Wording Blending 3 (Train)

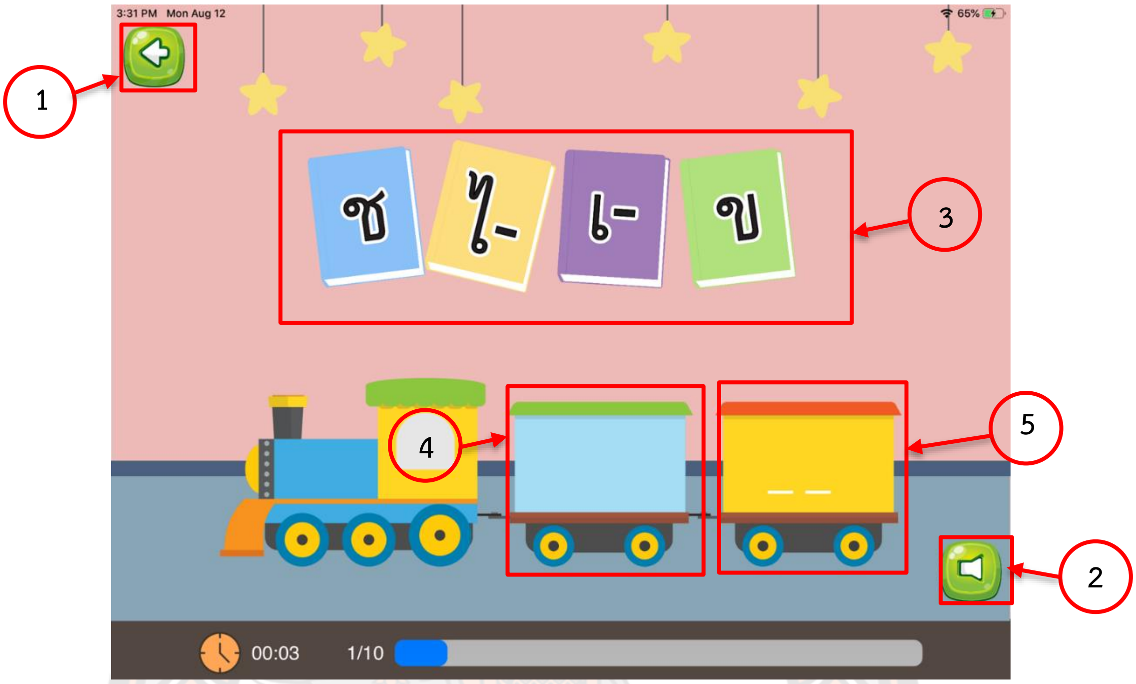


Figure 37. Screen of Word Blending 3 (Train)

1. Back to training menu button
2. Listen wording again
3. The game explains word before start “In this game, to give children look at the picture and listen carefully to the name of the picture. Then drag the correct answer and drop to the section 4”
4. Collect alphabets
5. Display the result word blending from the children

7. Review lessons menu

Review lessons divided into 6 lessons:

1. Initial Sound
2. Blending Alphabet + vowels (long tone)
3. Blending Alphabet + vowels (short tone)
4. Blending Alphabet + single vowels
5. Blending Alphabet + blending vowels
6. Blending Alphabet + excess vowels

Figure 38. The Review lessons menu screen

Number 1 Back to main menu button
Number 2 Select lesson to learn button

8. Recording of test results

8.1. If quit before the end of the game, the information must be saved first by pressing the OK button (number 1). To resume playing, press the Cancel button (number 2).

Figure 39. Confirm popup to exit and save

8.2. When end of the game, the system will be record data and give 1 heart

9. The result of child ability skill

The child's abilities are divided into 3 levels: letter sound, word blending easy level (alphabet + vowel) and word blending hard level (alphabet + vowel + ending), respectively.

9.1. Field the answer to shown result

Figure 40. Confirm popup to see the result

1. close popup
2. Field answer
3. OK button

Figure 41. Display the currently training status of child

Number 1: Back to main menu button

Number 2: The currently training status of child and show the syntax is passed and is fail or training

Institutional Review Board (IRB)

หนังสือรับรองการพิจารณาด้านจริยธรรมการวิจัยในคน

คณะอนุกรรมการจริยธรรมการวิจัยในคน มหาวิทยาลัยธรรมศาสตร์ ชุดที่ 1 (คณะแพทยศาสตร์)

95 หมู่ 8 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง จ. ปทุมธานี 12120

โทร. 02-9269704 , โทรสาร 02-5644444 ต่อ 7535

หนังสือรับรองเลขที่	089/2560
โครงการวิจัยเรื่อง	การทดสอบค่าความเที่ยงตรงและแม่นยำของเครื่องมือประเมินทักษะพื้นฐานของการอ่านภาษาไทยสำหรับเด็กชั้นประถมศึกษาปีที่ 1 ในรูปแบบโปรแกรมคอมพิวเตอร์ : Accuracy and Validity of the Computer Program for Assessment of Early Reading Skills: Thai Language in Grade 1 Students.
รหัสโครงการวิจัย	MTU-EC-PE-2-056/60
ผู้วิจัย	ผศ.พญ.อิสราภา ชื่นสุวรรณ
หน่วยงานที่รับผิดชอบ	โครงการจัดตั้งภาควิชากุมารเวชศาสตร์ คณะแพทยศาสตร์ มหาวิทยาลัยธรรมศาสตร์ โทร. 081-372-5032

เอกสารที่รับรอง

1. โครงร่างการวิจัย ฉบับที่ 2 วันที่ 11 พฤษภาคม 2560
2. เอกสารชี้แจงข้อมูลแก่ผู้เข้าร่วมโครงการวิจัย ฉบับที่ 2 วันที่ 11 พฤษภาคม 2560
3. หนังสือแสดงเจตนายินยอมเข้าร่วมการวิจัย ฉบับที่ 2 วันที่ 11 พฤษภาคม 2560
4. เอกสารข้อมูลสำหรับอาสาสมัครโครงการวิจัยสำหรับเด็กอายุ 7-<13 ปี ฉบับที่ 2 วันที่ 11 พฤษภาคม 2560
5. แบบสอบถาม ฉบับที่ 2 วันที่ 11 พฤษภาคม 2560

คณะอนุกรรมการจริยธรรมการวิจัยในคน มหาวิทยาลัยธรรมศาสตร์ ชุดที่ 1 (คณะแพทยศาสตร์) พิจารณาจริยธรรมการวิจัยโดยยึดหลักของ Declaration of Helsinki, The Belmont Report, CIOMS Guidelines และ the International Practice (ICH-GCP) ได้พิจารณาอนุมัติด้านจริยธรรมการทำวิจัยในคนให้ดำเนินการวิจัยข้างต้นได้ ตามมติที่การพิจารณาโครงการวิจัยแบบ Expedited Review

ระยะเวลาที่อนุมัติ 1 ปี

กำหนดส่งรายงานความก้าวหน้า 1 ปี : วันที่ 25 พฤษภาคม 2561

ลงชื่อ...
(รองศาสตราจารย์ นายแพทย์ไวพจน์ จันทร์วิเมลือง)
ประธานคณะอนุกรรมการฯ

ลงชื่อ...
(ผู้ช่วยศาสตราจารย์ ดร.สุมาลี คอนโด)
อนุกรรมการและผู้ช่วยเลขานุการ

อนุมัติ ณ **วันที่** 26 พฤษภาคม 2560
หมดอายุ **วันที่** 25 พฤษภาคม 2561

GLOSSARY

Keyword	Description
IRT	Item Response Theory
ICC	Item Characteristic Curve
1PL	The one-parameter logistic model
2PL	The two-parameter logistic model
3PL	The three-parameter logistic model
MLE	Maximum Likelihood Estimation
CM	Corrected Item Mean Substitution
API	Application Programming Interface

REFERENCES

[1] American Psychiatric Association. Specific learning disorder, In: Diagnostic and Statistical Manual of Mental Disorders, 5th ed, *American Psyciatric Assocociation, Arlingotn*, pp.66, 2013.

[2] S.K. Katusic, R.C. Colligan, W.J. Barbaresi, et al. Incidence of reading disability in a population-based birth cohort, Rochester, *Minn. Mayo Clin Proc, 76:1081*, pp. 1976-1982, 2001.

[3] R. Skiada, E. Soroniati, A. Gardeli, D. Zissis, EasyLexia: A Mobile Application for Children with Learning Difficulties, *In Procedia Computer Science*, vol. 27, pp.218-228, 2014.

[4] R. Yampratoom Eds., Emergent Literacy in Thai Preschoolers: A Preliminary Study, *Developmental & Behavioral Pediatrics*, vol. 38, no. 6, July., pp. 395-400, 2017.

[5] The Office of Educational Technology. "What is personalized learning" Jan, 2017. [Online]. Available https://medium.com/personalizing-the-learning-experience-insights/what-is-personalized-learning-bc874799b6f. [Accessed Jan. 19, 2019]

[6] S. Chen and J. Zhang, Ability Assessment based on CAT in Adaptive Learning System. *International Workshop on Education Technology and Training & International Workshop on Geoscience and Remote Sensing*, pp 88-91, 2008.

[7] P. Brusilovsky and C. Peylo, Adaptive and Intelligent Web-based Educational Systems, *Artificial Intelligence in Education*, vo1. 13, Apr., pp. 159-172, 2003.

[8] D.H. Sleeman and J.S. Brown, eds., *Intelligent tutoring systems*, Academic Press, 1982.

[9] D. Sampson and C. Karagiannidis, Personalised Learning: Educational, Technological and Standardisation Perspective, *Interactive Educational Multimedia*, vol. 4, Apr., pp. 24-39, 2002.

[10] M. Maddalora, Personalized Learning Model Using Item Response Theory, *International Journal of Recent Technology and Engineering (IJRTE)*, Vol. 8, June., pp. 811-818, 2019.

[11] C. Ming Chen, H. Ming Lee, and Y. Chen, Personalized e-learning system using Item Response Theory, *Computers & Education*, vol. 44, pp. 237-255, 2005.

[12] B. Baker and H. Kim, *Item Response Theory Parameter Estimation Techniques Second Edition, Revised and Expanded*. United States: Taylor & Francis Group, 2004.

[13] M. Yang and T. Kao, Item response theory for measurement validity, *Shanghai Arch Psychiatry*, vol. 26, no. 3, Jun., pp. 171-177, 2014.

[14] B. Csapo and G. Molnar, Online Diagnostic Assessment in Support of Personalized Teaching and Learning: The eDia System, *Advancements in Technology-Based Assessment: Emerging Item Formats, Test Designs, and Data Sources*, Jul., 2019

[15] Y. L. P. Vego, J. C. G. Bolanos, and G. M. F. Nieto, Application of item response theory (IRT) for the generation of adaptive assessment in an introductory course on object-oriented programming, *Frontiers in Education Conference Proceedings*, Oct., 2012

[16] A. Yacob, N. Hj. Ali and eds, Personalized Learning: An Analysis Using Item

Response Theory, *International Scholarly and Scientific Research & Innovation*, vol. 8, no. 4, pp. 1107-1113, 2014.

[17] Y. Kustiyahningsih and A. Cahyani, Computerized Adaptive Test based on Item Response Theory in E-Learning System. *International Journal of Computer Applications*, vol. 81, no. 6, Nov., pp. 6-11, 2013.

[18] C. Ming Chen and C. Ju Chung, Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle, *Computer & Education*, vol. 51, pp. 624-645, 2008.

[19] S. Rungrat, A. Harfield and S. Charoensiriwath, Applying item response theory in adaptive tutoring systems for Thai language learners, *11th International Conference on Knowledge and Smart Technology (KST)*, pp. 67-71, 2019.

[20] S. Rungrat, A. Harfield and S. Charoensiriwath, M-Learning Platform for Assessment and Personalized Learning of Thai Language by Primary School Children, *ICIC Express Letters, Part B: Applications*, vol. 12, no. 4, Apr., pp. 307-316, 2021.

[21] Learning Disabilities and Disorders, "Learning disabilities," Jun, 2015. [Online]. Available https://www.ctdinstitute.org/sites/default/files/ file_attachments/learning-disabilities-and- disorders.pdf. [Accessed Jun. 7, 2018]

[22] K. Vibulpatanavong, D. Evans, Off to a good start: Enhancing reading skills in Thai children.*1st International Conference on Special Education (ICSE)*, 2015.

[23] R. Yampratoom Eds., Emergent Literacy in Thai Preschoolers: A Preliminary Study, *Developmental & Behavioral Pediatrics*, vol. 38, no. 6, July., pp. 395-400, 2017.

[24] I. Chunsuwan, N. Ruangdaraganon, K. Vibulpatanavong and eds. Using Kidarn Application to Assess Thai Early Reading Skills: Evaluating Validity and Reliability. *Asia Pacific Journal of Developmental Differences*, vol. 7, no. 2, Jul., pp. 265-280, 2020.

[25] Atchariya Darote, KengThai application. 2013 [Mobile]. Available https://apps.apple.com/th /app/kengthai/id697559254. [Accessed Jun. 7, 2019]

[26] E. Bartle, Personalised learning: an overview, A discussion paper prepared for Professor Joanne Wright, Deputy Vice Chancellor (Academic) for the Vice-Chancellor's Retreat, *Institute for Teaching and Learning Innovation*, Mar., 23rd & 24th, 2015.

[27] F.H Wang. Application of Componential IRT Model for Diagnostic Test in a Standard- conformant eLearning System, *the Sixth International Conference on Advanced Learning Technologies* (ICAL'06), 2006.

[28] C. Ming Chen, Personalized curriculum sequencing utilizing modified item response theory for web-based instruction, *Expert Systems with Applications*, vol. 30, pp. 378–396, 2006.

[29] C. Ming Chen and C. Ju Chung, Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle, *Computer & Education*, vol. 51, pp. 624-645. 2008.

[30] K. Wauters, P. Desmet and W. Van den Noortgate, Adaptive item- based learning environments based on the item response theory: possibilities and challenges, *Computer Assisted Learning*, vol. 26, no. 6, pp. 549-562, Dec., 2010.

[31] Thai Language Institute Bureau of Academic affairs and Educational Standards

Office of the Basic Education Commission, *A guide to teaching reading and writing to spelling words*, Bangkok: The Agricultural Cooperative Federation of Thailand. Limited, 2018.

[32] B. Baker, *The Basics of Item Response Theory*, United States: ERIC Clearinghouse on Assessment and Evaluation, 2001.

[33] F. Holmes, Estimation of Item Response Theory Parameters in the Presence of Missing Data,. *Journal of Educational Measurement*, vol. 45, no. 3, pp. 225-245, 2008.

[34] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, Facing the cold start problem in recommender systems, *Expert Systems with Applications*, vol. 11, pp. 2065-2073, 2014.

[35] J Tendeiro, "IRT (GMMSGE01): Parametric IRT (dichotomous data)" Dec, 2017. [Online]. Available https://bookdown.org/jorgetendeiro/ParametricIRT. [Accessed Aug. 19, 2019]

BIOGRAPHY

Name-Surname	Saowalak Rungrat
Date of Birth	
Address	
Education Background	A.D. 2016 Computer Science, Naresuan University A.D. 2012 Nakhonsawan school
Awards	Semi Final Round the nineteen National Software Contest: NSC 2017