On Using Markov Chain to Evidence the Learning Structures and Difficulty Levels of One Digit Multiplication

Behnam Taraghi, Martin Ebner, Anna Saranti, Martin Schön
Graz University of Technology
Münzgrabenstrasse 35/I, 8010 Graz, Austria
{b.taraghi, martin.ebner}@tugraz.at, s0473056@sbox.tugraz.at, mrtnschn@googlemail.com

ABSTRACT
Understanding the behavior of learners within learning applications and analyzing the factors that may influence the learning process play a key role in designing and optimizing learning applications. In this work we focus on a specific application named "1x1 trainer" that has been designed for primary school children to learn one digit multiplications. We investigate the database of learners' answers to the asked questions (N > 440000) by applying the Markov chains. We want to understand whether the learners' answers to the already asked questions can affect the way they will answer the subsequent asked questions and if so, to what extent. Through our analysis we first identify the most difficult and easiest multiplications for the target learners by observing the probabilities of the different answer types. Next we try to identify influential structures in the history of learners' answers considering the Markov chain of different orders. The results are used to identify pupils who have difficulties with multiplications very soon (after couple of steps) and to optimize the way questions are asked for each pupil individually.

Categories and Subject Descriptors
H.2.8 [**Database Management**]: Database Applications – Data mining

General Terms
Algorithms, Measurement.

Keywords
Learning Analytics, Markov chain, difficulty level, one digit Multiplication, Math, elearning, primary school

LAK '14, March 24 - 28 2014, Indianapolis, IN, USA

http://dx.doi.org/10.1145/2567574.2567614

1. INTRODUCTION
The goal of Data Mining is to enhance the knowledge by interpreting implicit and gathered data. If the data itself or the context where it is used relates to education, the research approach is called Educational Data Mining [16]. Romero & Ventura stated in 2010 that educational data mining (EDM) is a *field that exploits statistical, machine-learning, and data-mining (DM) algorithms over the different types of educational data*. Learning Analytics (LA) can be seen as a further development and as a step towards a more human based assistance. Baker et al [1] mentioned that both research fields have an extensive overlap, but also subtle differences. EDM is more or less about using data to understand how learning occurs and how to improve it. LA strives to assist the learning process by giving educators a deeper insight to it based on data analyses. Duval [7] pointed out that we have to think about learners' traces and their learning efforts. Siemens and Baker [18] defined LA as *the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs*. So it can be concluded that the educator or the teacher plays an essential role in LA, because she or he is responsible for intervening in a pedagogical manner if the data analyses points out any hint [8].

Graz University of Technology has been developing math trainers since 2010 with the aim to improve the basic math education for primary schools [9]. First of all the so-called 1x1 trainer [17] was implemented, followed by the multi-math-coach [10] as well as the addition / subtraction trainer[1].

In primary schools, learning the one digit multiplication table is one of the major goals in the four-year period of education. Language implications in general [13], the role of math as first non-native language [14] and pure "row learning" [11] are some of the difficulties of this learning problem. Therefore a web-based application was developed which can both assist the learning process of the pupils and the pedagogical intervention of the teachers. According to the needs of the learners, four main parts are provided from the application: (a) the system is able to define a competence level of the learner; (b) the system is able to choose the given exercises according to the competence level of the learner. The implanted algorithm is responsible that the exercise is neither too easy nor too difficult; (c) the system ensures that already well-done exercises are repeated and practiced on a regular basis; (d) the system has to be appropriate for children from the age of 7-10 from the usability point of view. The full

[1] http://mathe.tugraz.at (last visit 24.01.2014)

implementation as well as the intelligent algorithm and the first results are described in [9]. This application was introduced to schools more than 1.5 years ago and it seems to be a success, because up to now, we have been able to gather more than 400.000 calculations done by children.

Several educational, pedagogical and psychological surveys classify various pupils' common errors in one digit multiplications and denote patterns of easy questions to learn, such as doubles, times five and square numbers [5, 12, 19]. However we could find no previous work dealing with the problem of one digit multiplication table computationally. This paper presents a computational analytical approach using a Markov chain model and classification algorithms.

In this publication we start first by analyzing the reaction times and the probabilities of the answer types. The goal is to detect the set of questions where the pupils mostly have problems to answer correctly. Next we go through studying the Markov chain model of different orders. Markov chains are used to model stochastic processes such as navigation models [3, 4, 6, 15]. We are especially interested in studying the regularities in learners' answers and finding the structures that may reveal a positive or negative effect on learners' answers to the subsequent questions. Finally we represent our results and discuss our findings.

2. METHODOLOGY – MARKOV CHAIN

A finite discrete Markov chain of the first order is a sequence of random variables $X_1, X_2, X_3, \ldots, X_n$ for which the following Markov property holds:

$$P(X_{n+1} = x_{n+1} \mid X_1 = x_1, X_2 = x_2, X_n = x_n) = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

We assume that the probabilities do not change as a function of time, hence the Markov chains are time-homogeneous.

A Markov chain of order k is described formally as follows:

$$P(X_{n+1} = {}_{n+1} \mid X_1 = x_1, X_2 = x_2, X_n = x_n) = P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1, \ldots,} X_{n-k+1} = x_{n-k+1})$$

The Markov chain of first order is called memoryless, meaning that the next state depends only on the current state. Considering the Markov chain of order k, the probability of the next state depends on the k previous states. A transition matrix P of all stochastic transition probabilities between the states represents the Markov model.

3. DATASET

In this paper we perform our analysis on the dataset from the database of "1x1 trainer" application. Our dataset initially contained 442910 answered questions by 3381 pupils over 11711 sessions. A first manual analysis of data made clear that there was some noise in the dataset such as unserious user answers. There were also cases where the users had left the application before reaching timeout. The dataset is cleaned and hence reduced to 438368 answered questions by 3377 pupils over 11609 sessions. The application puts each question to the pupils at least two times. Based on the correctness / incorrectness of the submitted answers in the user history, the answers for each question are classified into one of six different answer types. If a pupil answers a question for the first time correctly (answer type R that stands for RIGHT), the same question is asked once again later on to ensure that the answer is truly known by the pupil. If a question is answered correctly for the second time (answer type RR), the application assumes that the user had already known the answer to the put question truly. By contrast, if the pupil answers the question incorrectly in the second round (answer type RW), the application keeps on asking the same question later on until the pupil answers it correctly (answer type WR). Answer type W (stands for WRONG) implies the first incorrect answer to a question. Answer type WW implies an incorrect answer to a question after a preceding wrong answer. The following example illustrates how the answer types are assigned to each given answer. Assuming the application has put a pupil 5 times the question 9*3 in his history and the pupil's answers have been as follows: 27, 24, 26, 27, 27. The assigned answer types for this set of answers would be: R, RW, WW, WR, RR. The defined answer types build the states of the Markov chain model in our analysis. Table 1 lists these six defined answer types and their definitions.

Table 1. Six different answer types and their definition. "R" stands for "Right" and "W" for "Wrong".

Answer type	Definition	Preceding answer	Current answer
R	First correct answer	-	R
W	First wrong answer	-	W
RR	Correct answer given a preceding correct answer	R	R
RW	Wrong answer given a preceding correct answer	R	W
WR	Correct answer given a preceding wrong answer	W	R
WW	Wrong answer given a preceding wrong answer	W	W

Since our experiments are based on the statistical analysis, the distribution of the asked questions by the application may affect the reliability of the results. About 12 out of 90 questions exhibit a much lower frequency than the average. That is due to the fact that many pupils do not go through all the questions and leave the application too early. Consequently some of the questions are answered less frequently than the others.

4. DIFFICULT QUESTIONS

The probabilities of the occurring answer types in the dataset reveal the most difficult questions for the pupils. To identify the difficult questions most efficiently, we divided the dataset into two subsets. The first subset includes only the R and W answer types. The second subset includes only RW, WR and WW answer types. The following two subsections deal with these analyses in detail.

4.1 Subset: R and W Answer Types

This subset (R and W answer types only) includes the questions that are answered by the pupils for the first time either correctly or incorrectly. The goal is to identify the questions that were mostly already known (hence easy) und those that were mostly already unknown (hence difficult) to the pupils before a learning process actually begins. Figure 1 illustrates the histogram of the 30 most unknown asked questions (W answer type) within this subset sorted according to the frequency of unknown questions in descending order. At first sight, it is obvious that the pupils knew most of the questions beforehand. A total of 12531 out of 158741 questions were answered incorrectly. 7*8 and 6*8 are the first two questions with which most pupils had difficulties, whereas the frequency of both questions (unknown) is too close to each other.

As mentioned before, about 12 out of 90 questions exhibit a much lower frequency than the average in the dataset. Especially these questions are unapparent in the statistics, as a low number of pupils had answered. To overcome this problem, the proportional percentage of unknown (W answer type) to known (R answer type) questions was taken into consideration. Comparing the results, we were able to observe that the set of most unknown questions is mostly the same.

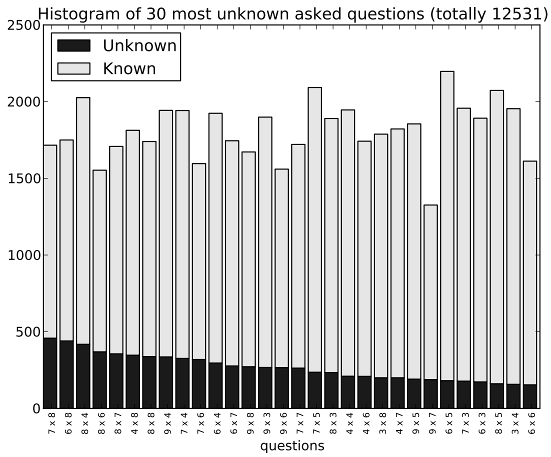

Figure 1. Histogram of the most 30 unknown asked questions within the subset of R and W answer types.

The "1x1 trainer" application provides ninety questions beginning with 1*1 to 10*9. Figure 2 shows a heatmap that illustrates these questions in respect of their easiness (known by high number of pupils) and difficulty (unknown by high number of pupils). As expected, it is to some extent symmetric. It illustrates that the multiplications where 1, 2, 5, and 10 occur as operand can be classified as easy or most known questions, whereas 3, 4, 6, 7, 8 and 9 as operands build multiplications that can be classified as difficult or most unknown. Considering the proportional percentage of questions instead of the pure number of unknown questions, the same results can be observed.

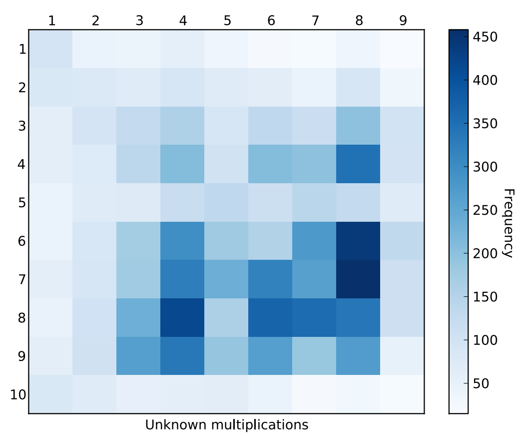

Figure 2. Heatmap of the asked unknown questions (multiplications) within the subset of R and W answer types. Rows and columns correspond to the multiplication operands.

If we consider the reaction times consumed by the pupils especially for the set of known (easy) answers (R answer types), we can observe that the pupils needed more time for the identified (difficult) unknown questions than the identified easy ones. Figure 3 illustrates this result in a heatmap. It shows the average time consumption for each question individually from the set of known questions. This observation confirms our results on the identified difficult and easy questions. Furthermore it is apparent that the pupils needed more time for 6*8 than 7*8. Our observation from the analysis of the second subset of data (subset of RW, WW and WR answer types) confirms our finding at this step (see the next subsection).

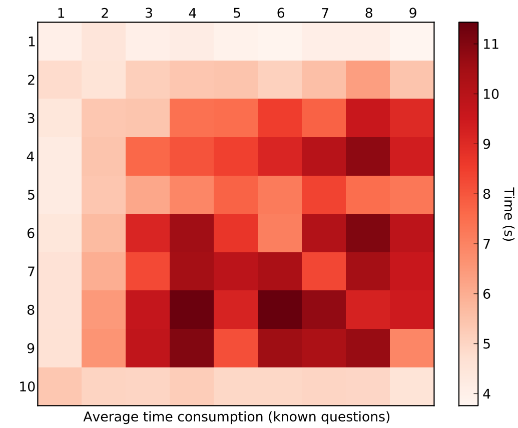

Figure 3. Heatmap of the average time consumption from the set of asked known questions (multiplications) within the subset of R and W answer types. Rows and columns correspond to the multiplication operands.

4.2 Subset: RW, WR and WW Answer Types

This subset (RW, WR and WW answer types) includes the questions answered by the pupils at least for the second time. The subset includes questions that the pupils answered incorrectly at least once in their history. The goal is to identify the questions with the highest probabilities (most difficult). These are the questions that the pupils repeated most often until they got the correct answer. Correspondingly the lowest probabilities reveal the questions that are easy to learn.

Figure 4 illustrates the histogram of the 30 most difficult questions within this subset.

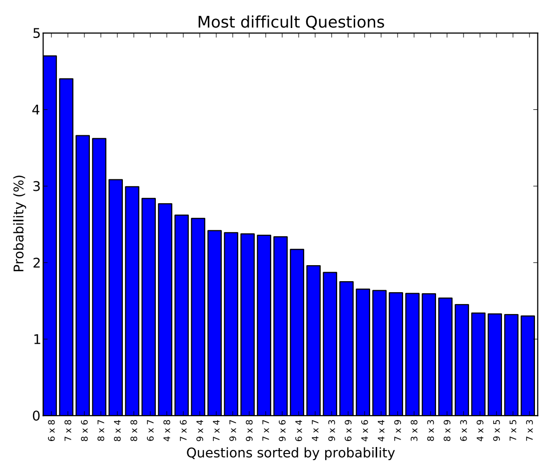

Figure 4. Histogram of the most 30 difficult questions within the subset of RW, WR and WW answer types.

6*8 and 7*8 are again the first two questions that have the highest probabilities. The heatmap of easy and difficult questions (low and high probabilities) is pretty much the same as the one from the subset of R and W answer types depicted in figure 2. The same operands were identified for easy and difficult questions. We divided the subset further into correct (WR answer types) and incorrect (RW, WW answer types) answers. We could observe the same results from both subsets as well. Considering the reaction times within this subset for correct answers (WR answer types) we could observe the same results as we did in the first subset (R and W answer types). The heatmap of the average time consumption for each question individually from the set of correct answers (WR answer type) looked pretty much the same as the set of R answer types depicted in figure 3. This observation confirms our findings about the identified difficult and easy questions.

5. STRUCTURES IN ANSWER TYPES

In this section we present our experimental results from the analysis of Markov chains. In our Markov model, the answer types to each question represent the states and the probabilities to the answer types of the subsequent questions in the sequence as transition links between the states. Hence we built a Markov matrix of the size 6*6 representing the transition probabilities between our 6 different defined answer types.

First we analyze the Markov model globally over all datasets. Figure 5 illustrates the Markov chain probabilities (in %) of self-transitions for each answer type by growing order k. These are transitions from one state to itself. For the sake of demonstration, only the first 5 orders are depicted here.

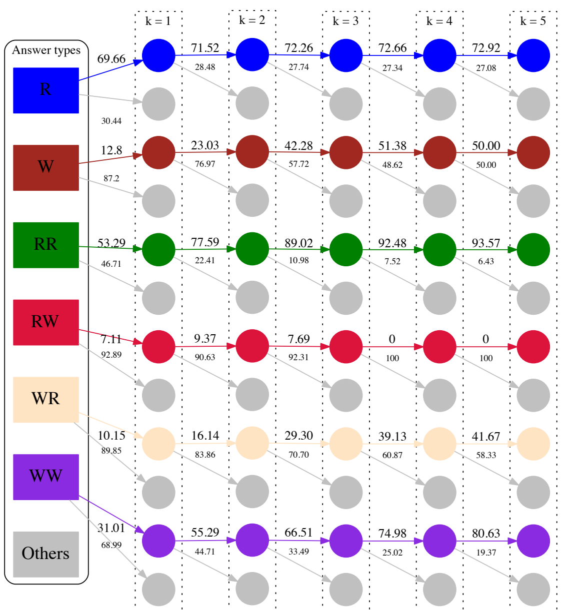

Figure 5. Markov chain probabilities of self-transitions (in %) for each answer type and for order k <= 5.

As can be seen, besides RW, the probabilities for all answer types increase by ascending k. For example the probability of transition from WW to WW state is about 31% (k = 1). By k = 2 the probability increases to 55.29%, by k = 3 to 66.51% and so forth. It means that assuming a pupil's last answer is of type WW, he will answer the next asked question with the probability of 31% incorrectly again (at least for the second time). If a pupil's last two answers are of type WW, the probability that he repeats this behavior increases to over 50% for the next asked questions. The k-probabilities for answer type R show that pupils who already know a question, will most probably (about 70%) know the next questions too. This is also true for RR answer type. In this case the probabilities increase even up to 93% in the fifth step. By contrast, we cannot claim the same for unknown questions. Our probability matrix shows than even for k is greater than 5, the probabilities for answer type W stay under 50%. The probabilities of answer type RW decreases after the second step und reaches 0 in the early forth step. This implies that if pupils once answer the questions correctly, they rarely give incorrect answers consecutively in the second round. The probability decreases as expected in the next steps to 0. The opposite effect can be seen for answer type WR. The probabilities increase slowly up to a maximum of 50% for k values greater than 5.

Another interesting observation we made relates to the alternative transitions between RW and WR states and vice versa. Figure 6 illustrates this structure clearly. Starting from the state RW, the transition to the WR state has the highest probability of 46.02%:

P (RW => WR) = 46% for k = 1

By k = 2 the highest probability belongs to the transition that switches back to the RW state:

P (WR => RW) = 61.02% for k = 2

By k = 3 the probability of transitions back to WR increases to 76.88% and so forth.

P (RW => WR) = 76.88% for k = 3

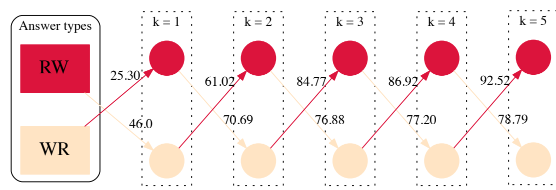

Figure 6. Alternative transitions between RW ⇔ WR states

This observation implies that the pupils who run into this structure would never come out of a loop switching between correct and incorrect answers for the same given questions. The same holds true for alternative transitions between WR and RW states. It can be observed very early, in particular from the second step (k = 1). In the case that pupils run into this structure while they answer the questions, teachers can be warned and made aware of that particular situation.

As a next step we investigated whether our outcomes can be observed in different difficulty levels relating to the questions. To reach this goal we used k-means algorithm [2] to classify the questions in three clusters corresponding to easy, intermediate and difficult levels. We gained 8 difficult, 22 intermediate and 60 easy questions. We repeated our experiment in each difficulty level individually. We could still observe increasing probability in self-transitions as described in the preceding subsection. However the alternative structure RW ⇔ WR could not be observed within the big set of easy questions.

6. DISCUSSION

In this work we analyzed the dataset from "1x1 trainer" application that was designed for primary school children to learn one-digit multiplications. We identified the easiest and the most difficult questions by looking through the probabilities of different answer types of the pupils in two different subsets. The reaction time of the pupils for answering the questions confirmed our results. The questions that were already known to the pupils before using the application correspond to the questions that were identified as easy through our analysis. The multiplications where 1, 2, 5, and 10 occur as operand can be classified as easy to learn, whereas 3, 4, 6, 9, and especially 7, 8 operands build multiplications that can be classified as difficult. The identified class of difficult questions contains eight multiplications as follows: 6*8, 7*8, 8*6, 8*7, 8*4, 8*8, 6*7 and 4*8. As can be seen, the difficult set is characterized mainly by the operand 8.

Next we analyzed the dataset by applying the Markov chain of higher orders and found some structures within pupils' answer types that can be relevant for predicting how the pupil will answer the forthcoming questions. Especially for difficult and intermediate questions it is of interest to teachers to know beforehand whether the pupils will have difficulties during work with the application. The goal is to support teachers to discover pupils' weak points during training in a fast and accurate way.

7. CONCLUSION

The goal of our research was to develop applications for basic math education, which allow individualizing learning. Each child should be assisted in their own and personal way. Therefore a detailed overview was built to assist teachers in their daily work. The data analyses presented in this publication will help to improve the application in two different ways: the current empirical estimated difficulties of the question have to be adapted to the outcomes of this research work and the estimated learning state has to be improved according to the outcomes of the analysis of Markov chains.

Future research studies will be carried out if the two measurements help pupils learn the multiplication table in a more individualized way. Bearing in mind that in the past, teachers had never been able to get such exact data about the learning process of their pupils, and how to intervene in an appropriate pedagogical manner, learning analytics must be seen as an important enabler of future education.

8. REFERENCES

[1] Baker, R. S. et al. .2012. Panel: Educational Data Mining meets Learning Analytics. LAK 12: 2nd International Conference on Learning Analytics & Knowledge

[2] Bishop, C. 2006. *Pattern Recognition and Machine Learning*. Springer Science and Business Media, LLC, New York, USA, 424-430

[3] Borges, J. and Levene, M. 2007. Evaluating variable-length Markov chain models for analysis of user web navigation sessions. IEEE Trans. On Knowl. And Data Eng., 19(4): 441-452.

[4] Brin, S. and Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. In Proceedings of the seventh international conference on World Wide Web 7, WWW7. Pages 107-117. Amsterdam, The Netherlands. Elsevier Science Publishers B. V.

[5] Chambers, D. 1996. Direct modeling and invented procedures: Building on children's informal strategies. Teaching Children Mathematics, 3, 92-95.

[6] Deshpande, M. and Karypis G. 2004. Selective Markov models for predicting web page accesses. ACM Trans. Internet Technol., 4(2): 163-184.

[7] Duval, E. 2010. Attention Please! Learning Analytics for Visualization and Recommendation. In proceedings of LAK11: 1st International Conference on Learning Analytics and Knowledge 2011.

[8] Ebner, M., Neuhold, B. and Schön, S. 2013. Learning Analytics - wie Datenanalyse helfen kann, das Lernen gezielt zu verbessern. In K. Wilbers & A. Hohenstein (Hrsg.), Handbuch E-Learning. Expertenwissen aus Wissenschaft und Praxis – Strategien, Instrumente, Fallstudien. Köln: Deutscher Wirtschaftsdienst (Wolters Kluwer Deutschland), 48. Erg.-Lfg. Jänner 2013. pp 1-20

[9] Ebner, M. and Schön, M. 2013. Why Learning Analytics in Primary Education Matters! Bulletin of the Technical Committee on Learning Technology, Karagiannidis, C. & Graf, S (Ed.), Volume 15, Issue 2, April 2013, p. 14-17

[10] Ebner, M., Schön, M., Taraghi, B. and Steyrer, M. 2013. Teachers Little Helper: Multi-Math-Coach, Proceedings of the IADIS International Conference e-Learning 2013, Nunes, M. B. & McPherson, M. (Ed.), Prague, IADIS Press, p. 183-190

[11] Fuson, K.C. 1990. Using a base-ten blocks learning/teaching approach for first- and second-grade- place-value and multidigit addition and subtraction . In: *Journal for Research in Mathematics Education*. 21, No. 3, 180-206 Northwestern University . DOI= http://webschoolpro.com/home/projectlead/Research%20Articles%20and%20links/Base%20Ten%20Blocks.pdf

[12] Garnett, K. 1992. Developing fluency with basic number facts: Intervention for students with learning disabilities. Learning Disabilities Research & Practice 7, 210-216.

[13] Gerster, H.D. 2009. Schwierigkeiten bei der Entwicklung arithmetischer Konzepte im Zahlenraum bis 100.- In: *Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie*. Fritz, A., Ricken, G., Schmidt, S. (Ed.)- Weinheim, Basel, Berlin: Beltz, p 248–268.

[14] Landerl, K., Butterworth, B., Lorenz, J.H., and Möderl, K. 2003. Rechenschwäche - Rechenstörung - Dyskalkulie: Erkennung - Prävention – Förderung. Leykam, Graz.

[15] Lempel, R. and Moran, S. 2000. The stochastic approach for link-structure analysis (salsa) and the tkc effect. Comput. Netw., 33(1-6):387-401.

[16] Romero, C. and Ventura, S. 2010. Educational Data Mining: A Review of the State of the Art. IEEE Transactions on systems, man and Cybernetics – Part C: Applications and Reviews, Vol. 40, No. 6., 601- 618

[17] Schön, M., Ebner, M. and Kothmeier, G. 2012. It's Just About Learning the Multiplication Table, In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12), Simon Buckingham Shum, Dragan Gasevic, and Rebecca Ferguson (Eds.). ACM, New York, NY, USA, 73-81

[18] Siemens, G. and Baker, R. S. J. 2012. Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. LAK '12 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pp. 252-254.

[19] Thornton, C. 1990. Strategies for the basic facts. In J. Payne (Ed.), Mathematics for the young child, pp. 133-51, Reston, VA: National Council of Teachers of Mathematics.