Towards Personalizing An E-quiz Bank for Primary School Students: An Exploration with Association Rule Mining and Clustering

Xiao Hu
Faculty of Education,
University of Hong Kong
Pokfulam, Hong Kong
xiaoxhu@hku.hk

Yinfei Zhang
Faculty of Education,
University of Hong Kong
Pokfulam, Hong Kong
zhangyinfei99@gmail.com

Samuel K. W. Chu
Faculty of Education,
University of Hong Kong
Pokfulam, Hong Kong
samchu@hku.hk

Xiaobo Ke
School of Management
Guangdong University of Technology
1833724585@qq.com

ABSTRACT
Given the importance of reading proficiency and habits for young students, an online e-quiz bank, Reading Battle, was launched in 2014 to facilitate reading improvement for primary-school students. With more than ten thousand questions in both English and Chinese, the system has attracted nearly five thousand learners who have made about half a million question answering records. In an effort towards delivering personalized learning experience to the learners, this study aims to discover potentially useful knowledge from learners' reading and question answering records in the Reading Battle system, by applying association rule mining and clustering analysis. The results show that learners could be grouped into three clusters based on their self-reported reading habits. The rules mined from different learner clusters can be used to develop personalized recommendations to the learners. Implications of the results on evaluating and further improving the Reading Battle system are also discussed.

Categories and Subject Descriptors
• Information systems → Association rules • Information systems → Clustering • Applied computing → E-learning

Keywords
Association rule mining; clustering; e-quiz bank; reading.

1. INTRODUCTION
Reading proficiency is fundamental for students as it is closely associated with their learning abilities and academic performance [2]. It is widely recognized that the advancement of reading abilities is essential to primary-school students as it facilitates their understanding of academic materials and tasks [1], [2]. As a result, students with higher reading proficiency tend to perform better in academic activities [1], [2].

Reading Battle (http://battle.cite.hku.hk/) is an online e-quiz system designed to develop primary-school students' reading interests and enhance their reading proficiency [2]. Launched in 2014, Reading Battle has included over 500 English and Chinese books available in local school libraries, as well as more than 10,500 well-tailored questions related to these books. The books are categorized into a 23-category taxonomy (e.g., "fiction" or "folklore"), in English and Chinese (see Table 4). Each question in the system is also encoded with a question category (ranging from "information search", "interpretation and synthesis", "inference" to "evaluation") and a difficulty level (ranging from level 1 to level 4). The materials are cataloged and developed by reading literacy experts including language teachers and school librarians, with the guidance of the international PIRLS assessment framework on children's reading literacy [2].

After reading a book registered to the system, a student can log in to the Reading Battle system and take a quiz corresponding to that book. Around 30 questions are designed for each book and each time the system randomly selects 10 questions to form a quiz for the student. The questions in Reading Battle are in the same language (English or Chinese) as that of the books they are associated with. The bilingual setting is uniquely designed, following the policy of "two written languages and three spoken codes" in Hong Kong. To date, the Reading Battle system has attracted over 30 primary schools in Hong Kong and Taiwan, a public library summer program and a kindergarten in the U.S., with a total of over four thousand student users.

A significant trend in the education field is to deliver personalized learning experience to the learners. There has been strong empirical evidence on the positive effects of personalized learning on academic performance and learning effectiveness [3]. As there are different categories of books and questions, and students' interests and abilities vary, it would be more effective if the system can provide personalized learning experience by recommending books a student likes and questions that fit his or her abilities.

This study thus aims to discover potentially useful knowledge from reading and question answering records of primary-school students enrolled in Reading Battle. The analysis results will be used to improve the system design and develop personalized recommendations of books and questions to learners. As our first attempt along this direction, this paper aims to answer the following research questions:

1) What kinds of books and questions have the students read and answered?
2) Are there interesting association rules among the books the students read?
3) Are there interesting association rules among the questions the students answered?

Answers to the first question help develop overall understanding of the usage patterns of books and questions since the launch of

LAK '16, April 25 - 29, 2016, Edinburgh, United Kingdom

ACM 978-1-4503-4190-5/16/04…$15.00
DOI: http://dx.doi.org/10.1145/2883851.2883959

Reading Battle. The examination of association rules on the book level will reveal potentially preferable reading orders. An association rule on the book level would be like "if a student reads book X (or books in category A), he/she is likely to read book Y (or books in category B)". Such rules can be readily used to make book recommendations to students based on their reading records.

The examination of association rules on the question level will provide insights on the question design. If a student answers questions in category A (or level X) correctly, association rule mining will help find out whether this student is likely to answer questions in category B (or level Y) correctly. Such rules will help indicate whether the design of question categories and difficulty levels is suitable for the student users. Suggestions for system improvement can be made based on the indications. Furthermore, the results can be used to generate personalized question sets to accommodate individual student's abilities.

By analyzing learner's question answering records in Reading Battle, this study helps develop a comprehensive understanding on how primary-school students use an e-quiz bank for reading, particularly a bilingual system such as Reading Battle. It also provides evidence on how well the system is serving the student users, based on what suggestions for further improvement of the system can be proposed. Furthermore, the rules mined from learners' records can be used to deliver personalized book recommendations and question selection, which will likely be a significant enhancement to the learning experience. Overall, the study is expected to provide empirical evidence on how learning analytics can be applied to enhance the understanding of learners and improve the learning environment.

2. METHODS

2.1 Association Rule Mining and Clustering

Association rule mining and clustering are common data mining methods used to analyze user data from learning systems [6]. Association rule mining is used to explore the relationships between items in a large dataset. It helps discover rules in the form of "premise → conclusion" which stands for "if the premise occurs in the dataset, the conclusion is likely to occur as well". Such rules are very useful for disclosing relationships between items in a dataset. Frequent Pattern Growth (FP-Growth) algorithm is used for association rule mining in this study, for its efficiency and clarity as implemented in the RapidMiner toolkit [5]. Whether a resultant rule is sufficiently strong or interesting is decided by a series of measures. In this study, we follow the suggestions from Merceron and Yacef [7] on rule mining in the education domain, and employ the Cosine and Lift measures of interestingness to evaluate rules output by the algorithm, with the minimum Cosine threshold set as 0.65 and the minimum Lift as 1.10 [7]. Considering both measures helps ensure the retained rules are meaningful and interesting. Clustering is a data mining technique for discovering proximity patterns in a given dataset [4]. Data samples are separated into different groups such that a high similarity is expected within groups while different groups should be able to differentiate themselves from each other. Agglomerative clustering algorithm is adopted in this study for its deterministic nature and flexibility in deriving the number of resultant clusters.

2.2 Data Collection and Analysis

The data analyzed in this study is a snapshot of the Reading System taken in May 2015. There were a total of 3,175 students producing 26,189 book reading records and 457,235 question answering records in Reading Battle as of May 12, 2015. Among these active users, 523 of them responded to an online questionnaire about their reading behaviors, preferences and attitudes towards reading. As this group of learners and their parents have given consent to the project team to analyze the data, in this study we extracted and analyzed the 119,377 system records of these learners. Two sets of user data were extracted from the system: 1) students' reading records, each including book title, unique ID of the books, book category and language; 2) students' quiz records, each of which includes the question, question answer, question category and difficulty level. Clustering was applied to group the students into clusters based on their questionnaire answers. Association rule mining was applied to analyze the data on four levels: 1) book category; 2) individual book; 3) question difficulty level; and 4) question category.

3. RESULTS AND DISCUSSION

3.1 Student Clusters

A total 523 responses were received for the online survey and the demographic information is summarized in Table 1. All but five of the students were primary students ranging from Grade 1 to Grade 6. The majority of them were in Grade 5, accounting for 67.4% of the students considered in this study.

Table 1: Student demographics

Characteristics		N	%
Gender	Male	270	51.6
	Female	253	48.4
	Total	523	100.0
Grade	P1	3	0.6
	P2	15	2.9
	P3	56	10.8
	P4	81	15.6
	P5	349	67.4
	P6	14	2.7
	Missing	5	1.0
	Total	523	100.0

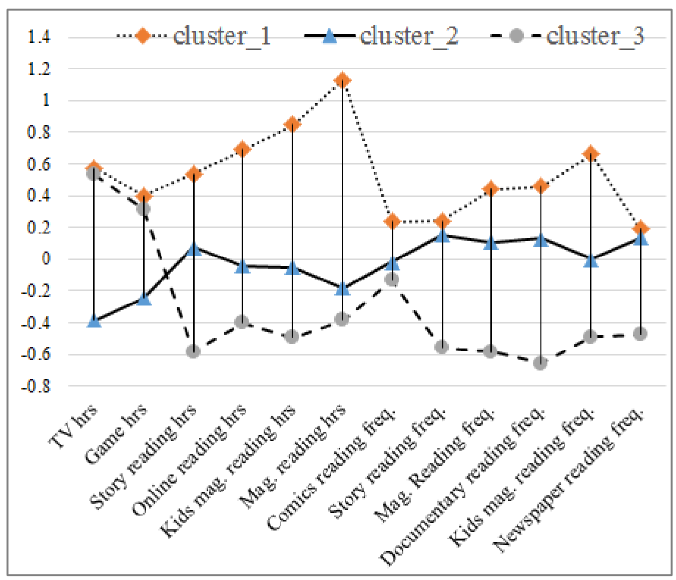

Figure 1: Centroid plot of three clusters

Cluster analysis was carried out to group the students into clusters based on their responses to 12 questions related to reading habits, including hours and frequencies of reading various kinds of books, as well as hours of watching TV and playing games. The agglomerative clustering results showed three major clusters and thus three clusters were formed. The normalized values of the cluster centroids are presented in Figure 1. Cluster 1 (n=91) consists of students with long reading hours and a high reading frequencies, and thus it is identified as *the cluster with good reading habits*. Cluster 3 (n=124) with short reading hours and low reading frequencies is identified as *the cluster with bad reading habits*. Cluster 2 (n=308) lies in between the other two clusters, with middle values of reading time and reading frequencies. It is thus identified as *the cluster with moderate reading habits*.

Association rule mining was then applied to mine interesting rules from the reading and question answering records among all students and students in each cluster. Table 2 provides an overview of the interesting and non-redundant rules for each student group on the book and question levels.

Table 2: Numbers of interesting and non-redundant association rules for each student cluster

	Book category	Individual book	Question category	Question level
All students	1	0	19	29
Cluster 1	46	16	43	77
Cluster 2	11	4	12	24
Cluster 3	2	0	15	20

3.2 Book-level Rules

3.2.1 Book Category

Twenty three book categories were defined for English and Chinese books in the Reading Battle. Table 3 presents the number of book in each category and how many times each category has been read. Picture book and fiction were among the most popular categories in both languages. In addition, English fantasy, Chinese fairytale and Chinese science book were also read quite frequently.

Table 3: Book categories

Book category	English Book		Chinese book	
	Number	Times being read	Number	Times being read
Biography	2	19	7	63
Detective story	2	15	9	137
Fable	7	101	1	36
Fairytale	9	71	26	283
Fantasy	38	114	2	18
Fiction	30	124	60	285
Folklore	6	28	21	227
Nonfiction	16	35	23	80
Picture book	57	163	17	221
Science	14	73	34	299
Science & picture book	1	25	0	0
Fantasy & picture book	0	0	1	43
History	0	0	4	104
Total	182	768	205	1,796

Based on the criteria described in Section 2.1, only one interesting rules was mined among all 523 students considered in our study. A closer look at candidate rules discloses their low Support values, which means most of the students read books in different categories. However, the Support values of candidate rules increase when student clusters are considered separately. This indicates students with similar reading habits (in the same cluster) tended to read books in the same categories more often than students with different reading habits (in different clusters). Table 4 shows all the interesting rules mined on book categories in records of students in Cluster 2 (*moderate reading habits*), two of which are also found among those mined from Cluster 3 (*bad habits*). In fact, Cluster 3 has very few interesting rules (n=2), probably due to the fact that those students did not read much and thus had few records in the system. Most of the rules in Table 4 are also included in rules mined from Cluster 1 (*good habits*) which in fact has much more rules (n=46) that cannot be enumerated in Table 4. In addition, Cluster 1 is the only cluster with rules involving English book categories. Not surprisingly, this is because students in this cluster read a lot and produced many records in the system. Table 5 presents a sample of interesting rules mined from Cluster 1 and their measures.

Table 4: Interesting and non-redundant rules on book categories for student clusters

No.	Premises	Conclusion	Cluster 1	Cluster 2	Cluster 3
1	童話(fairytale)	民間故事 (folklore)	√	√	-
2	小說 (fiction)	童話 (fairytale), 科學 (science)	√	√	-
3	民間故事 (folklore)	童話 (fairytale), 科學 (science)	√	√	-
4	民間故事 (folklore)	童話 (fairytale), 小說 (fiction)	√	√	-
5	童話 (fairytale), 科學 (science)	小說 (fiction), 民間故事 (folklore)	-	√	-
6	童話 (fairytale)	科學 (science), 小說 (fiction)	√	√	-
7	科學 (science)	圖畫故事 (picture book)	√	√	-
8	圖畫故事 (picture book)	科學 (science)	√	√	-
9	小說 (fiction)	民間故事 (folklore)	√	√	-
10	科學 (science)	童話 (fairytale), 小說 (fiction)	√	√	-
11	科學 (science)	民間故事 (folklore)	-	√	-
12	童話 (fairytale)	小說 (fiction)	√	√	√
13	小說 (fiction)	童話 (fairytale)	√	√	√

Recommendations can be made based on the rules for each cluster. For example, if a student with good or moderately good reading habits (Cluster 1 or 2) has read a Chinese fairytale, he/she is likely to read a Chinese folklore (Rule #1 in Table 4). Next time when he/she logs in to the system, a recommendation list of Chinese folklores can be presented on his/her homepage. In the case of Cluster 3 (*bad habits*), after exhausting the two rules of its own (i.e. recommending Chinese fiction to students who have read Chinese fairytale and vice versa), rules from Cluster 2 (*moderate habits*) can be borrowed and applied to the students in Cluster 3 (*bad habits*). The premise of this "rule borrowing" strategy lies in that rules in a student cluster with better reading habits can be beneficial to students in another cluster with worse reading habits, and that rules in one cluster can largely be acceptable by students in an adjacent cluster. Similarly, when rules in Cluster 2 (*moderate habits*) have all been applied to a student in Cluster 2, rules in Cluster 1 (*good habits*) can be applied to this student. As there are more rules in clusters with better reading habits, this strategy can help mitigate the potential problem of rule deficiency. Of course, the effectiveness of the strategy needs to be evaluated with students, which will be conducted in our future work.

Table 5: Sample of interesting rules on book categories for students with good reading habits (Cluster 1)

Premises	Conclusion	Support	Confidence	Lift	Cosine
Fable	小說(fiction) ,童話(fairytale), picture book, fiction	0.21	0.76	2.93	0.78
Fable	科學(science), picture book	0.21	0.76	2.67	0.74
Fable	小說(fiction), 圖畫故事(picture book), 童話(fairytale)	0.21	0.76	2.10	0.66
Fiction	小說(fiction), 童話(fairytale), 民間故事(folklore), picture book	0.22	0.59	2.38	0.72
圖畫故事 (picture book) , 童話(fairytale), fiction	民間故事(folklore), picture book	0.21	0.76	2.44	0.71

3.2.2 Individual Book

A total of 387 books were read by the students, among which 362 books were read more than one time. No interesting rules are discovered in the entire student population or among the students in Cluster 3. A few interesting and non-redundant rules are mined for Cluster 2 (n=4) and Cluster 1 (n=16). The reason for the small number of rules is that there are many individual books in the system and reading records of individual books are sparse. Limited recommendations of individual books can be made based on the rules from these student clusters. Cluster 3 can then borrow the rules from the other two clusters considering that the other two clusters have better reading habits and more interesting rules. More reading records are needed to establish interesting and strong rules on the individual book level.

3.3 Question-level Rules

There were 4,436 and 5, 968 English and Chinese questions that have been answered by the students respectively. Their distribution across difficulty levels and categories is shown in Table 6. As in books, there were more Chinese questions answered than English ones in all difficulty levels and categories except for "Evaluation".

Table 6: Question difficulty level and category

	English Questions	Chinese Questions
Difficulty level		
Level 1	1,540	1,577
Level 2	1,439	1,756
Level 3	648	1,461
Level 4	809	1,174
Total	4,436	5,968
Category		
Information search (infn)	2,184	2,139
Interpretation & synthesis (inte)	781	2,036
Inference (infer)	901	1,386
Evaluation (eva)	570	407
Total	4,436	5,968

Interesting association rules mined from all students are listed in Table 7, Rule 1 to 16 on question difficulty levels and Rule 17 to 27 on question categories. It is not surprising that answering questions of higher difficulty levels correctly would imply answering those of lower difficulty levels correctly, for both English and Chinese questions (Rule #4, 8, 9 10, 11, 12). However, among Chinese questions, quite a few rules also indicate the opposite: answering questions of lower difficulty levels correctly could imply answering those of higher difficulty levels correctly (Rule # 1, 2, 3, 5, 6, 7). The high Support and Confidence values of Rule #1 and #2 indicate these cases occurred quite often. In contrast, rules of English questions seem to follow the difficulty levels well. No rules with correct answers in lower difficulty levels alone imply correct answers in higher difficulty levels. This seems to indicate that the difficulty levels of English questions are better scaled and more suitable for the student users in consideration. For Chinese questions, Rule #1 and #3 are symmetric (i.e., answering Level 1 questions correctly implies answering Level 2 questions correctly, and vice versa), so are Rule #2 and #3. These rules show that, to this group of students, the difficulty levels of Chinese questions in Level 1 and Level 2, and those in Level 1 and Level 3 may not be clearly distinguishable. One possible reason is that Chinese questions of Level 1-3 are relatively easy for this group of student users. A closer look at the rules from student clusters further verifies this conjecture: all three clusters contain rules like Rule # 1 and 2, as well as [level_2_Chi_r] → [level_3_Chi_r]. Therefore, it can be suggested to the Reading Battle system to increase the difficulty degree of Level 2 and Level 3 Chinese questions, so that they would be more challenging to the students.

The case is a little different for Level 4 Chinese questions. To have the conclusion of answering Level 4 Chinese questions correctly, the premise clauses have to include correctly answering questions in two other levels at the same time, i.e. Level 1 and Level 2 questions (Rule #5), Level 1 and Level 3 questions (Rule #6), or Level 2 and Level 3 questions (Rule #7). These rules indicate that Level 4 Chinese questions are indeed more demanding than those in other levels. Rule #14 to #16 reveal associations among questions across languages. It seems that correctly answering questions in a higher difficulty level in one language could help boost the chance of correctly answering questions in a lower difficulty level in the other language. However, this needs to be verified when more data are collected and more rules can be mined.

For question categories, Rule #17 and #18 indicate if students answered "Evaluation" Chinese questions correctly, they would likely to answer all other three kinds of Chinese questions correctly. For these students, it is recommended that they proceed to a higher difficulty level in Chinese questions in order to keep them challenged and continuing improving their abilities. For English questions, Rule #24 to #26 indicate if students correctly answered any other kind of English questions than "Information search", they would correctly answer English questions in "Information search" category. These results are in accordance with the common knowledge that "Evaluation" questions are of a higher complexity than others and "Information search" questions are more straightforward than others. Rule #19 to #21 reveal two symmetric pairs: (infer_Chi_r, infn_Chi_r) and (infer_Chi_r, inte_Chi_r). Another symmetric pair between (infn_Chi_r, inte_Chi_r) is revealed by Rule # 22 and 23. These pairs are also found in each of the three student clusters. They suggest that the complexity levels of Chinese questions in these three categories (infn, inte, and infer) are mostly comparable, to the students considered in this study. Again, similar to difficulty levels, the results suggest that Reading Battle probably should adjust the complexity of Chinese questions in these three categories, so that they can be more useful in differentiating students' abilities in information search, interpretation & synthesis and inference.

Table 7: Interesting and non-redundant rules among question difficulty levels and categories for the entire student sample

No.	Premises	Conclusion	Support	Confidence	Lift	Cosine
Question difficulty level						
1	level_1_Chi_r	level_2_Chi_r	0.66	0.87	1.20	0.89
2	level_1_Chi_r	level_3_Chi_r	0.64	0.85	1.16	0.86
3	level_2_Chi_r	level_1_Chi_r, level_3_Chi_r	0.59	0.82	1.27	0.87
4	level_3_Chi_r	level_1_Chi_r, level_2_Chi_r	0.59	0.81	1.22	0.85
5	level_1_Chi_r, level_2_Chi_r	level_4_Chi_r	0.55	0.83	1.32	0.85
6	level_1_Chi_r, level_3_Chi_r	level_4_Chi_r	0.54	0.84	1.33	0.85
7	level_3_Chi_r, level_2_Chi_r	level_4_Chi_r	0.53	0.83	1.31	0.83
8	level_4_Chi_r	level_1_Chi_r, level_3_Chi_r, level_2_Chi_r	0.51	0.81	1.37	0.84
9	level_2_Eng_r	level_1_Eng_r	0.33	0.85	1.97	0.81
10	level_3_Eng_r	level_1_Eng_r	0.30	0.86	2.00	0.77
11	level_4_Eng_r	level_1_Eng_r, level_2_Eng_r	0.25	0.80	2.40	0.77
12	level_4_Eng_r	level_3_Eng_r	0.25	0.80	2.32	0.75
13	level_1_Eng_r, level_3_Eng_r	level_2_Eng_r	0.25	0.84	2.14	0.73
14	level_3_Chi_r, level_1_Eng_r	level_2_Eng_r	0.26	0.80	2.05	0.73
15	level_4_Chi_r, level_1_Eng_r	level_2_Eng_r	0.24	0.81	2.05	0.70
16	level_2_Chi_r, level_3_Eng_r	level_4_Chi_r, level_1_Eng_r	0.22	0.81	2.72	0.77
Question category						
17	eva_Chi_r	infn_Chi_r, inte_Chi_r	0.50	0.85	1.24	0.79
18	eva_Chi_r	infn_Chi_r, infer_Chi_r	0.48	0.81	1.27	0.78
19	infer_Chi_r	infn_Chi_r, inte_Chi_r	0.60	0.86	1.26	0.87
20	infn_Chi_r	infer_Chi_r	0.64	0.84	1.20	0.87
21	inte_Chi_r	infer_Chi_r	0.63	0.84	1.20	0.87
22	inte_Chi_r	infn_Chi_r	0.68	0.91	1.19	0.90
23	infn_Chi_r	inte_Chi_r	0.68	0.90	1.19	0.90
24	infer_Eng_r	infn_Eng_r	0.31	0.88	2.02	0.79
25	inte_Eng_r	infn_Eng_r	0.29	0.87	2.01	0.76
26	eva_Eng_r	infn_Eng_r	0.28	0.90	2.07	0.77
27	infn_Eng_r, inte_Eng_r	infer_Eng_r	0.23	0.81	2.31	0.73

Note: infn=information search; inte=interpretation & synthesis; infer=inference; eva=evaluation; Chi=Chinese; Eng=English; r=right/correct.

4. CONCLUSION AND FUTURE WORK

The study explores the associations among books and questions in an e-quiz system, Reading Battle, which supports primary-school students in improving their reading interests and proficiency. Clustering and association rule mining techniques were used to analyze user records extracted from the system. The resultant association rules can be used to improve the question design and develop personalized recommendations to student users. The results also help further deepen our understanding of students' usage of the Reading Battle system, which in turn provides rationales for further improvement. As the system continues to run, more user data will be collected in the future. The expansion of record size for each student will not only enlarge the scale of association rules but also help improve the strength and confidence of the rules. Besides, participation in the online survey of reading habits should also be encouraged in the future, so that more students can be grouped into clusters and cluster-based recommendations could be made available to them. In addition, the online survey provides an opportunity to collect relevant data for building up user profiles based on which more rules for recommendation can be explored.

This study strives to provide empirical evidence that learning analytics can be applied to mine valuable knowledge from student interaction with online learning system and facilitate the development of personalized learning, particularly for an online reading platform for primary-school students. It contributes to the literature on the development of personalization in online reading platforms. As the next step, we will implement the rules identified in this study into the recommender system in Reading Battle and evaluate their effectiveness with real users.

REFERENCES

[1] Baker, L. and Wigfield, A. 1999. Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly*, 34, 4, 452-477. DOI= http://dx.doi.org/10.1598/RRQ.34.4.4

[2] Chu, S., Chan, H., Wong, J., Wu, W., and Mok, S. 2014. Strengthening students' reading comprehension ability (both Chinese and English) through developing children's literature equiz bank on the cloud. *The 19th International Education and Technology Conference*, Hong Kong. DOI= http://hdl.handle.net/10722/204504

[3] Duran, J. I., Laitakari, J., Pakkala, D., and Perala, J. 2010. A user meta-model for context-aware recommender systems. In *Proceedings of the HetRec,* Barcelona, Spain. DOI= http://dx.doi.org/10.1145/1869446.1869456

[4] Han, J., Kamber, M., and Pei, J. 2011. *Data mining: concepts and techniques*. Elsevier.

[5] Hofmann, M., and Klinkenberg, R. 2013. (Eds.). *Rapidminer: Data Mining Use Cases and Business Analytics Applications*. CRC Press.

[6] Hsu, C.-K., Hwang, G.-J., and Chang, C.-K. 2010. Development of a reading material recommendation system based on a knowledge engineering approach. *Computers & Education*, 55, 1, 76-83. DOI= http://dx.doi.org/10.1016/j.compedu.2009.12.004

[7] Merceron, A. and Yacef, K. 2008. Interestingness measures for association rules in educational data. *EDM, 57-66.*